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

AVALIAÇÃO DO DESEMPENHO DOS ALUNOS DA PRIMEIRA SÉRIE DO COLÉGIO TÉCNICO INDUSTRIAL DE SANTA MARIA

MONOGRAFIA DE ESPECIALIZAÇÃO

Mauricio Ramos Lutz

Santa Maria, RS, Brasil

2006

AVALIAÇÃO DO DESEMPENHO DOS ALUNOS DA PRIMEIRA SÉRIE DO COLÉGIO TÉCNICO INDUSTRIAL DE SANTA MARIA

por

Mauricio Ramos Lutz

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

AVALIAÇÃO DO DESEMPENHO DOS ALUNOS DA PRIMEIRA SÉRIE DO COLÉGIO TÉCNICO INDUSTRIAL DE SANTA MARIA

elaborada por
Mauricio Ramos Lutz

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)

Lindolfo Storck, Dr. (UFSM)

Santa Maria, 10 de março de 2006.

O essencial é saber ver,
Saber ver sem estar a pensar,
Saber ver quando se vê,
E nem pensar quando se vê
Nem ver quando se pensa
Mas isto (tristes de nós que trazemos
A alma vestida!)
Isso exige um estudo profundo,
Uma aprendizagem do desaprender

Fernando Pessoa
(Alberto Caeiro)

AGRADECIMENTOS

À professora Luciane Flores Jacobi, por ter me acompanhado durante esta trajetória de construção deste trabalho, entendendo minhas experiências, fornecendo subsídios teóricos e orientando a produção desta monografia, sorrindo e motivando sempre, e por dedicar tempo à leitura e análise deste trabalho;

Ao meu pai, in memoriam;

À minha mãe, que motivou, acompanhou e contribuiu, de todas as formas possíveis e, de modo especial, emocionalmente para a realização deste trabalho, amo você;

Aos amigos e companheiros Magda, Rogério, Rejane, Alícia, Mara, Luiz, Carla, Moacir, enfim a todos, que escutavam as preocupações de estudante, acompanhando as alegrias e dificuldades vivenciadas durante a realização do curso;

Aos professores e colegas da turma que partilharam momentos de formação teórica, de amizade e de acolhida;

Acima de tudo, "que eu jamais esqueça de Deus e de que um pequeno grão de alegria e esperança dentro de cada um é capaz de mudar e transformar qualquer coisa, pois a vida se concretiza no amor." (autor desconhecido)

Muito obrigado a todos!!!

RESUMO

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

AVALIAÇÃO DO DESEMPENHO DOS ALUNOS DA PRIMEIRA SÉRIE DO COLÉGIO TÉCNICO INDUSTRIAL DE SANTA MARIA
AUTOR: MAURICIO RAMOS LUTZ
ORIENTADORA: LUCIANE FLORES JACOBI
Data e local da Defesa: Santa Maria, 10 de março de 2006.

Estabelecer relações, encontrar ou propor leis explicativas é o papel próprio da ciência. A análise multivariada corresponde a um grande número de métodos e técnicas que utilizam simultaneamente todas as variáveis na interpretação teórica do conjunto de dados obtidos, sendo o método escolhido para trabalhar com os dados, a análise de medidas repetidas. Os dados coletados pertencem a 61 alunos matriculados regularmente na primeira série do Ensino Médio do Colégio Técnico Industrial de Santa Maria e analisados pelo programa computacional *Statistica 7.0*, com a finalidade de verificar se a origem educacional influencia nas notas, se o desempenho escolar entre os gêneros são equivalentes e, também, investigar se as disciplinas, pelas notas médias formariam grupos. Realizada esta análise, verificou-se que não há diferenciação entre os tipos de escola, mas sim entre algumas disciplinas, e, também, que as alunas possuem um desempenho significativamente superior ao dos alunos, em todas as disciplinas independentes da escola de origem.

Palavras-chave: Estatística, Análise multivariada, Análise de dados repetidos, Desempenho escolar.

ABSTRACT

Specialization Monograph
Post Graduated Program in Statistics and Quantitative Modeling
Federal University of Santa Maria

EVALUATION OF THE PERFORMANCE OF THE FIRST GRADE STUDENTS OF THE FIRST SERIES OF SANTA MARIA'S INDUSTRIAL TECHNICAL SCHOOL
AUTHOR: MAURICIO RAMOS LUTZ
ADVISOR: LUCIANE FLORES JACOBI
Date and place of defense: Santa Maria, March 10th , 2006.

To establish relationships, to find or to propose explanatory laws is the role of science itself. The Multivariate Analysis corresponds to a great number of methods and techniques that simultaneously use all the variables in the theoretical interpretation of the group of colleted data. The chosen method to work with the data is the analysis of repeated measures. The collected data comes from 61 students registered regularly in the first grade of the high school of Santa Maria's Industrial Technical School, and was analyzed by the Statistica 7.0 computer program, with the purpose of verifying how educational origin influences the grades, if the study performance among the genders is equivalent and also to investigate if the subjects, by the mean grades, would form groups. The results of this analysis show that there is no difference among the kinds of school, but between some subjects, and also that the girls get better grades than the boys no matter the school of origin or subject.

Word-keys: Statistics, Multivariate analysis, Repeated data analysis, study performance.

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE SIGLAS

B – Biologia

EA – Educação Artística

EF – Educação Física

F – Estatística da distribuição F calculada

FI – Física

GL – Grau de Liberdade

LB – Literatura Brasileira

LI – Língua Inglesa

LP – Língua Portuguesa

M – Matemática

P – Projetos

p – valor da probabilidade exata do teste

Q – Química

QM – Quadrado Médio

SQ – Soma dos Quadrados

LISTA DE ANEXOS

SUMÁRIO

1 INTRODUÇÃO

Em qualquer decisão que tomemos em nossas vidas, sempre levamos em conta um grande número de fatores. Obviamente nem todos pesam da mesma maneira na hora de uma escolha. Às vezes, por tomarmos uma decisão usando a intuição, não encontramos de maneira sistemática estes fatores, ou seja, não identificamos quais as variáveis que afetam a decisão.

Quando analisamos o mundo que nos cerca, reconhecemos que todos os acontecimentos, sejam eles culturais ou naturais, envolvem um grande número de variáveis. As diversas ciências têm o desejo de conhecer a realidade e de interpretar os acontecimentos (ciências humanas) e os fenômenos (ciências naturais), baseados no conhecimento das variáveis intervenientes consideradas importantes nesses eventos.

A exigência da mão-de-obra qualificada, de profissionais competentes para atuar no mundo globalizado e competitivo requer um constante aperfeiçoamento. Estabelecer relações, encontrar ou propor leis explicativas é o papel próprio da ciência. Para isso é necessário controlar, manipular, medir as variáveis que são consideradas relevantes ao entendimento do fenômeno analisado. Muitas são as dificuldades em traduzir as informações obtidas em conhecimento. A maior delas é de natureza epistemológica: a ciência não conhece a realidade, apenas a representa através de modelos e teorias dos diversos ramos do conhecimento. Outra dificuldade é a aspiração de universalidade das explicações científicas. Ora, isto implica e condiciona a pesquisa a uma padronização metodológica. Um aspecto essencial desta padronização é a avaliação estatística das informações.

O desenvolvimento tecnológico oriundo das descobertas científicas tem alavancado o próprio desenvolvimento científico, ampliando em várias ordens de grandeza a capacidade de obter informações de acontecimentos e fenômenos que são analisados. Uma grande massa de informação deve ser processada antes de ser transformada em conhecimento. Portanto, cada vez mais necessitamos de ferramentas estatísticas que apresentem uma visão global do fenômeno que aquela possível em uma abordagem univariada. A denominação análise multivariada corresponde a um grande número de métodos e técnicas que utilizam simultaneamente todas as variáveis na interpretação teórica do conjunto de dados obtidos.

Johnson e Wichern (1998) afirma que dentre os métodos de análise multivariada temos a análise de componentes principais, análise de correspondência, análise de agrupamentos e análise de medidas repetidas. O método utilizado para trabalhar com os dados em estudo é a análise de medidas repetidas, logo tendo um foco para seu procedimento.

A necessidade de uma ferramenta para a demonstração do desempenho comparativo entre as diferentes escolas de origem, gêneros e disciplinas faz-se necessária para uma melhor avaliação destes alunos, objetivando uma melhor preparação por parte da escola e também facilitando a tomada de decisões.

1.1 Objetivos

1.1.1 Objetivo Geral

O presente trabalho tem por objetivo avaliar o desempenho escolar dos alunos da 1ª série do Ensino Médio do Colégio Técnico Industrial de Santa Maria.

1.1.2 Objetivos Específicos

- Verificar se a origem educacional influencia nas notas;
- Comparar o desempenho escolar entre os gêneros;
- Investigar a existência de grupos de disciplinas com afinidade.

1.2 Justificativa

Conforme Hair Jr (1998), análise multivariada corresponde a um grande número de métodos e técnicas que utilizam simultaneamente todas as variáveis na interpretação teórica do conjunto de dados obtidos, pondo em evidência ligações, semelhanças ou diferenças, aliados a estes motivos expostos, justificas-se a utilização desta análise.

O que realmente justifica esta temática abordada é a verificação no âmbito escolar em saber se existe ou não diferença significativa no desempenho dos alunos nas várias disciplinas em relação à origem educacional e ao gênero, para um melhor planejamento pedagógico e tomada de decisões.

1.3 Metodologia

Para o desenvolvimento do presente trabalho, utilizou-se os dados de 61 alunos matriculados regularmente na primeira série do Ensino Médio no ano de 2004 do Colégio Técnico Industrial de Santa Maria, localizado na Universidade Federal de Santa Maria. Os dados foram fornecidos pela secretaria da escola e constavam do nome do aluno, data, local e estado de nascimento, gênero, idade, tipo de escola de origem, estado e ano de conclusão do Ensino Fundamental e das referidas notas obtidas na primeira série do Ensino Médio. Foram analisadas 10 disciplinas, entre elas, Literatura Brasileira, Língua Portuguesa, Língua Inglesa, Educação Artística, Educação Física, Física, Matemática, Química, Biologia e Projetos.

Após a coleta dos dados, realizou-se a análise de variância multivariada para medidas repetidas e quando necessário a comparação de médias pelo teste de Tukey, utilizando, para tanto, o programa computacional *Statistica 7.0*.

1.4 Estrutura do trabalho

Esta pesquisa realizada divide-se em 4 capítulos. No Capítulo 1, tem-se a introdução, os objetivos, geral e específicos, a justificativa, a metodologia e a estrutura do trabalho; no Capítulo 2, apresentação da revisão de literatura sobre a análise de medidas repetidas; no Capítulo 3, os resultados e discussões e no Capítulo 4, conclusão.

2 REVISÃO DE LITERATURA

A análise multivariada de medidas repetidas é um método de análise estatística que permite um estudo global das variáveis, pondo em evidência ligações, semelhanças ou diferenças.

Segundo Singer, Rocha e Nobre (2004) entende-se por medidas repetidas como sendo os dados gerados através de observações de um número de unidades de investigação repetidamente sobre diferentes condições de avaliação, assumindo-se que as unidades de investigação constituem uma amostra aleatória de uma população de interesse.

De uma forma geral, este tipo de pesquisa envolve observações de um conjunto de unidades de investigação (animais, por exemplo) classificados em diferentes subpopulações segundo um ou mais fatores, ou tratamentos (como raça, procedência, tipo de ração), ao longo de diversas condições de avaliação (como o tempo, dose) que representam as unidades de observação.

Estudos com medidas repetidas envolvem a investigação da evolução de uma ou mais variáveis, respostas observadas em diferentes unidades amostrais através do tempo ou de outra escala ordenada.

Conforme Andrade e Singer (1986) por meio do número de condições de avaliação (ocasiões) em que cada unidade de investigação é observada, pode-se reconhecer três categorias principais de planejamentos:

i) transversal (*cross-sectional*), onde se planeja observar cada unidade de investigação em uma única ocasião;

ii) longitudinal, onde se planeja observar as unidades de investigação durante determinados intervalos de tempo. Esses intervalos podem ser comuns a todas as unidades de investigação ou não;

iii) misto, onde se planejam observar diferentes grupos de unidades de investigação em diferentes intervalos de tempo.

Os planejamentos longitudinais têm duas motivações importantes:

i) produzem estimativas mais precisas do efeito do fator que define as condições de avaliação, que um planejamento transversal com o mesmo número de observações, (realizadas em distintas unidades de investigação);

ii) são especialmente indicados para o estudo de mudanças individuais ao longo do tempo.

Mais especificamente, dados longitudinais podem ser definidos como medidas repetidas onde as observações dentro das unidades de investigação não foram ou não podem ter sido atribuídas aleatoriamente para as diferentes condições de avaliação (usualmente tempo ou posição no espaço).

Na prática 3 (três) tipos de análise estatística são comumente usados na análise de medidas repetidas.

i) Análise de variância univariadas como se fosse um experimento em parcela subdividida (*spit-plot*) com tratamentos como fator de parcela e o tempo como o fator da subparcela;

ii) Análises univariada e multivariada para transformações lineares das medidas repetidas, tais como inclinações e outras tendências em curvas de regressão, divergentes entre respostas de diferentes pontos de tempo;

iii) Métodos baseados em métodos 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 3 métodos considerados:

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

ii) 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;

iii) Metodologia de modelos mistos que permite eficientemente conduzir a questão diretamente na modelagem da estrutura de covariância.

Stuker (1986) analisa as medidas repetidas, utilizando estas metodologias onde segue 2 etapas:

i) Modela a estrutura de covariância;

ii) Analisa as tendências de tempo para os tratamentos estimados e comparando médias.

Em situações de medidas repetidas, onde há uma estrutura balanceada e completa, tem-se uma configuração com grupos, unidade experimental e condições de avaliação, conforme a Tabela abaixo:

Tabela 1 – Estrutura de dados básicos para estudos com planejamento transversal.

Grupos de tratamentos	Unidade experimental	Condições de avaliação			
		1	2	...	p
1	1	y_{111}	y_{112}	...	y_{11p}
1	2	y_{121}	y_{122}	...	y_{12p}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
1	n_1	$y_{1n_1 1}$	$y_{1n_1 2}$	...	$y_{1n_1 p}$
2	1	y_{211}	y_{212}	...	y_{21p}
2	2	y_{221}	y_{222}	...	y_{22p}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
2	n_2	$y_{2n_2 1}$	$y_{2n_2 2}$	...	$y_{2n_2 p}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
g	1	y_{g11}	y_{g12}	...	y_{g1p}
g	2	y_{g21}	y_{g22}	...	y_{g2p}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
g	n_g	$y_{gn_g 1}$	$y_{gn_g 2}$	...	$y_{gn_g p}$

onde y_{ijk} é o valor da variável resposta da j-ésima unidade experimental dentro do i-ésimo tratamento, sob a k-ésima condição de avaliação (tempo, por exemplo) para $i = 1,2,..., g$; $j = 1,2,..., n_i$; $k = 1,2,..., p$ e $\sum_{i=1}^{g} n_i = N$.

Segundo Singer, Rocha e Nobre (2004), as técnicas clássicas de análise de dados longitudinais são geralmente dirigidas para o caso de dados completos e balanceados em relação ao tempo, destacando-se:

i) a análise multivariada e univariada de perfis que pode ser realizada através de modelos mistos;

ii) a análise de curvas de crescimento que pode ser realizada através de modelos uni e multivariados.

A análise univariada de perfis é uma das técnicas estatísticas utilizadas na análise de observações, provenientes de experimentos com medidas repetidas. Baseia-se no número total de observações e requer algumas suposições na estrutura da matriz de variância e covariâncias entre as observações.

Riboldi *et al*. (2002), considera que um conjunto de dados, com a estrutura apresentada na Tabela 1, a utilização da técnica de análise univariada de perfis exige a suposição de que as observações para cada unidade experimental sejam expressas em unidades comparáveis, obedeçam a uma distribuição normal com variância constante em cada uma das p condições de avaliação e covariância constante entre duas observações dentro de cada unidade experimental, e, também, que as unidades experimentais sejam independentes.

A análise univariada de perfis caracteriza-se pelo fato de que as comparações entre tratamentos podem ser divididas em dois tipos: comparações realizadas dentro da unidade experimental (ou seja, controlando as diferenças entre unidades experimentais) e comparações que são feitas entre unidades experimentais, isto é, envolvendo unidades experimentais diferentes. Este fato implica uma divisão de variação residual em dois tipos básicos: uma delas é a variação entre unidades experimentais (conhecida como variação entre sujeitos) e a outra é a variação dentro das unidades experimentais (conhecida como variação dentro de sujeitos). Estes dois tipos de variação são considerados separadamente na análise de variância e correspondem a medidas de variância entre e dentro de sujeitos (Riboldi *et al*., 2002).

Na análise multivariada os vetores de resposta, são independentes, seguem uma distribuição normal p-variada, com vetor de médias $\mu(px1)$ e matriz de variâncias e covariâncias, Σ, igual a:

$$\sum_{(pxp)} = \sigma^2 \begin{bmatrix} 1 & \rho & \rho & \cdots & \rho \\ \rho & 1 & \rho & \cdots & \rho \\ \rho & \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho & \rho & \rho & \cdots & 1 \end{bmatrix}$$

onde ρ é uma constante que representa a correlação entre dois elementos quaisquer dentro do mesmo vetor de respostas. Este tipo de matriz Σ é conhecida como matriz uniforme ou com padrão de uniformidade (Riboldi *et al*. ,2002).

Desta forma, um modelo matemático adequado para analisar dados longitudinais é o mesmo utilizado para analisar experimentos em parcelas subdivididas (*split-plot*). Neste modelo, consideram-se tratamentos e condições de avaliações como dois fatores cruzados fixos e as unidades experimentais (sujeitos) como um fator aleatório hierárquico a tratamentos (Riboldi *et al*. ,2002).

A utilização do modelo matemático é:

$$y_{ijk} = \mu + \alpha_i + \delta_{j(i)} + \beta_k + \alpha\beta_{ik} + \varepsilon_{ijk} \qquad (1)$$

onde μ representa a média geral populacional; α_i representa o efeito do i-ésimo tratamento; $\delta_{j(i)}$ representa o efeito aleatório de j-ésima unidade experimental dentro do i-ésimo tratamento; β_k representa o efeito da k-ésima condição de avaliação; $\alpha\beta_{ik}$ representa o efeito da interação entre o i-ésimo tratamento e a k-ésima condição de avaliação; ε_{ijk} representa o erro aleatório da observação y_{ijk}, onde

$$\delta_{j(i)} \sim IID\left(o, \sigma_\delta^2\right) \qquad (2)$$

$$\varepsilon_{ijk} \sim ID\left(0, \sigma_\varepsilon^2\right) \qquad (3)$$

$$E\left(y_{ijk}\right) = \mu + \alpha_i + \beta_k + \alpha\beta_{ik} \qquad (4)$$

$$V\left(y_{ijk}\right) = \Sigma \qquad (5)$$

Conforme Riboldi *et al.* (2002), a Tabela de análise de variância (ANOVA) é a mesma do modelo de Parcela Subdividida. Sendo esta equivalente a (MANOVA) que pode ser vista em Morrison (1990) e que é dado por:

Tabela 2 – Quadro da análise de variância para o modelo proposto.

Fonte de Variação	GL	SQ	QM	F
Tratamento	$g-1$	SQ_1	QM_1	F_1
Erro (a) – Unidade de investigação (dentro do tratamento)	$N-g$	SQ_2	QM_2	
Tempo ou condição	$p-1$	SQ_3	QM_3	F_2
Tratamento x tempo	$(g-1)(p-1)$	SQ_4	QM_4	F_3
Erro (b) – tempo x unidade de investigação (dentro de tratamento)	$(N-g)(p-1)$	SQ_5	QM_5	
Total	$Np-1$	SQ_6		

onde:

$$SQ_1 = \left(\frac{1}{N}\right)\sum_{i=1}^{g} y_{i..}^2 - \left(\frac{1}{Np}\right)y_{...}^2; \tag{6}$$

$$SQ_2 = \left(\frac{1}{p}\right)\sum_{i=1}^{g}\sum_{j=1}^{r} y_{ij.}^2 - \left(\frac{1}{N}\right)\sum_{i=1}^{g} y_{i..}^2; \tag{7}$$

$$SQ_3 = \left(\frac{1}{N}\right)\sum_{k=1}^{p} y_{..k}^2 - \left(\frac{1}{Np}\right)y_{...}^2; \tag{8}$$

$$SQ_4 = \left(\frac{1}{p}\right)\sum_{i=1}^{g}\sum_{k=1}^{p} y_{i.k}^2 - \left(\frac{1}{Np}\right)y_{...} - SQ_1 - SQ_3; \tag{9}$$

$$SQ_5 = SQ_6 - SQ_1 - SQ_2 - SQ_3 - SQ_4; \tag{10}$$

$$SQ_6 = \sum_{i=1}^{g}\sum_{i=1}^{r}\sum_{i=1}^{p} y_{ijk}^2 - \left(\frac{1}{Np}\right)y_{...}^2; \tag{11}$$

$$QM_1 = \frac{SQ_1}{g-1}; \tag{12}$$

$$QM_2 = \frac{SQ_2}{N-g}; \tag{13}$$

$$QM_3 = \frac{SQ_3}{p-1}; \tag{14}$$

$$QM_4 = \frac{SQ_4}{(g-1)(p-1)}; \tag{15}$$

$$QM_5 = \frac{SQ_5}{(N-g)(p-1)}; \tag{16}$$

$$F_1 = \frac{QM_1}{QM_2}; \tag{17}$$

$$F_2 = \frac{QM_3}{QM_5}; \tag{18}$$

$$F_3 = \frac{QM_4}{QM_5}. \tag{19}$$

Nas expressões acima, o uso de ponto (.) no índice indica soma com respeito àquele índice e de barra (–) sobre y indica média aritmética. Por exemplo:

$$\bar{y}_{i..} = \left(\frac{1}{N}\right)y_{i..} = \left(\frac{1}{N}\right)\sum_{j=1}^{r}\sum_{k=1}^{p} y_{ijk}. \tag{20}$$

Os testes univariados para efeitos dentro de sujeitos (tempo) e da interação envolvendo efeitos entre e dentro de sujeitos (tratamentos x tempo) requerem algumas suposições para que as probabilidades provenientes do teste F usual estejam corretas.

Posto que a análise de variância multivariada observa pressuposições generalizadas decorrentes da análise de variância univariada, as observações devem seguir uma distribuição normal multivariada, isto é, os dados devem constituir amostras aleatórias oriundos de populações normais, e também haver igualdade de matrizes de variância e covariância, ou seja, essas amostras deverão pertencer a grupos populacionais com idêntica variância. (Hair, 1998).

Com base nas propriedades da distribuição normal, Reis (1997) afirma que qualquer combinação linear de variáveis normais é também normal e que todos os contornos da função distribuição normal multivariada formam uma elipse.

Não existe nenhum teste, de acordo com a literatura, que permita uma análise de um conjunto de dados que segue ou não uma distribuição normal e conjunta para as variáveis. Assim, o que se faz é testar cada variável em separado para a normalidade, mesmo tendo certeza de que as variáveis permaneçam normais quando verificadas separadamente, supomos que também mantenham a normalidade em tamanhos mais elevados.

Segundo Reis (1997), para a verificação do pressuposto de normalidade temos que seguir três condições observadas nos dados. A primeira é que as distribuições marginais para o conjunto de observações são normalmente distribuídas. Em segundo, as representações gráficas de pares de observações representam uma elipse e, em terceiro, não existam observações muito diferentes (outliers), ou seja, que seja resultado apenas de erros de introdução de dados.

As representações gráficas, especificamente, o gráfico Q-Q é uma alternativa para se verificar a distribuição normal de determinada população; este gráfico pode ser utilizado para avaliar se determinada distribuição marginal segue a normalidade. Os quartis da distribuição amostral e os correspondentes valores esperados, são os pares de valores representados por este tipo de gráfico (Reis, 1997).

O conceito de distância de Mahalanobis entre dois vetores de observações, é usado para o caso multivariado. Podendo ser calculada da seguinte maneira:

$$d_j^2 = \left(X_i - \overline{X}\right)^T S^{-1}\left(X_j - \overline{X}\right), \qquad (21)$$

onde T significa transposto.

Para a determinação do gráfico Q-Q, deve-se seguir quatro passos:

1º Passo: ordenar as distâncias $d^2_{(j)}$ por ordem crescente;

2º Passo: calcular as probabilidades acumuladas $p_{(j)} = \dfrac{j-0,5}{n_j}$ para cada um dos valores anteriores;

3º Passo: definir o percentil de ordem 100. $p_{(j)}$ na distribuição χ^2 com p graus de liberdade;

4º Passo: representar no gráfico os pares ordenados $\left(d^2_{(j)}, \chi^2_{\mathrm{p}(j)}\right)$.

Após a construção do gráfico Q-Q, observa-se se os pares ordenados $\left(d^2_{(j)}, \chi^2_{\mathrm{p}(j)}\right)$ se aproximam de uma linha reta, mais aproximados estarão de uma distribuição normal multivariada; caso contrário, se os dados possuírem um afastamento do comportamento linear, sugerem que os dados não apresentam uma distribuição normal multivariada (Reis 1997).

Verifica-se, atendido o pressuposto de normalidade multivariada, a igualdade das matrizes de variância e covariância.

O teste M desenvolvido por Box (1950 apud REIS, 1997) é uma generalização do teste univariado de igualdade de variância de Bartlett. Observando-se que esse teste é sensível a violações do pressuposto de normalidade, atribuído ao fato de utilizar as variâncias generalizadas, isto é, os determinantes das matrizes de covariância. Posto isto, a hipótese de igualdade de matrizes pode ser rejeitada apenas pela violação do pressuposto de normalidade e não por se tratarem de matrizes significativamente diferentes.

O autor ainda define o teste M utilizando o método do quociente de verossimilhanças, assumindo que os vetores de médias dos grupos são desconhecidos, sendo que se deve testar as seguintes hipóteses:

$$H_o : \Sigma_1 = \Sigma_2 = ... = \Sigma_i \text{ com } \hat{\Sigma} = \frac{T}{N-1} = S \text{ e } \hat{\mu}_j = \overline{X}_j$$

$$H_1 : \Sigma_i \neq \Sigma_j \quad i \neq j \text{ com } \hat{\Sigma}_i = S_i \text{ e } \hat{\mu}_j = \overline{X}_j ,$$

onde:

N a dimensão total da amostra;

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:

$$M = (N - i)\ln|S_c| - \sum_{I=1}^{i} v_i \ln|S_i| \qquad (22)$$

onde:

$N = \sum_{I=1}^{i} n_i$ dimensão total da amostra; (23)

$v_i = n_i - 1$ os graus de liberdade associados a cada grupo; (24)

$S_c = \frac{1}{\sum n_i} \sum_{I=1}^{i} n_i S_i$ é a matriz de variância e covariância comum; (25)

i é o número de grupos de tratamento.

Box (1950 apud REIS, 1997, p. 186), sugeriu duas aproximações para o teste M. A primeira é uma aproximação pela distribuição χ^2; dada por:

$$M.C \sim \chi^2_{\left[\frac{1}{2}k(k+1)(i-1)\right]} \quad , \qquad (26)$$

onde:

$C = 1 - \frac{2k^2 + 3k - 1}{6(k+1)(i-1)} \left(\sum_{I=1}^{i} \frac{1}{v_i} - \frac{1}{N-1} \right)$ para grupos com tamanhos de amostras diferentes. (27)

E, $C = 1 - \frac{(2k^2 + 3k - 1)(i+1)}{6(k+1)iN}$ para grupos com tratamentos de amostras iguais. (28)

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

E a segunda maneira é feita pela aproximação da distribuição F, dada por:

$$\frac{M\left(1 - a_1 - \frac{v}{v_0}\right)}{v} \sim \mathrm{F}_{(\mathrm{v},\mathrm{v}_0)} \qquad (29)$$

onde:

$a_1 = 1 - C$ (30)

$a_2 = \frac{(k-1)(k+2)}{6(i-1)} \left(\sum_{I=1}^{i} \frac{1}{v_i^2} - \frac{1}{(N-1)^2} \right)$ (31)

$v = \frac{k(k+1)(i-1)}{2}$ (32)

e $v_0 = \dfrac{v+2}{a_2 + a_1^2}$ (33)

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

A análise de variância tem por objetivo principal apontar se um grupo é ou não estatisticamente diferente. Caso a hipótese nula (H_0), seja rejeitada a um determinado nível de significância, existirá pelo menos uma das médias de um tratamento que difere da(s) outra(s). Na determinação de qual ou quais tratamentos não são iguais estatisticamente, é utilizado a diferença mínima significativa (d.m.s.), que serve para comparar as médias dos tratamentos (Vieira, 1999).

Segundo Vieira (1999), uma das formas para se encontrar a d.m.s é o teste de Tukey, que permite realizar comparações entre médias duas a duas, verificando-se, assim, se existe ou não diferença significativa entre as médias de um grupo.

Para a realização do teste de Tukey, primeiramente encontra-se a d.m.s dado por:

$$dms = q_{\alpha;(\delta,k)} \sqrt{\frac{QMR}{r}}, \qquad (34)$$

onde:

q é o valor tabelado, levando-se em consideração os graus de liberdade do resíduo (δ), o número de tratamentos (k) e o nível de significância (α);

QMR é o quadrado médio dos resíduos;

r é o número de repetições de cada tratamento.

Sempre que o valor absoluto da diferença entre duas médias é igual ou maior do que o valor da d.m.s., diz-se que as médias são estatisticamente diferentes, ou seja, se $\left|\overline{X}_i - \overline{X}_j\right| \geq d.m.s.$, para $i \neq j$ (Vieira, 1999).

A análise multivariada de perfis é uma das técnicas estatísticas utilizadas para analisar observações provenientes de experimentos com medidas repetidas. Esta técnica se baseia no número de unidades experimentais como tamanho da amostra, enquanto o procedimento univariado de análise de perfis utiliza, como tamanho de amostra, o número total de observações. 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. Ambos procedimentos, univariado e multivariado, exigem a suposição de que as componentes aleatórias do erro devem seguir uma distribuição normal (Singer, Rocha e Nobre 2004).

Uma desvantagem do procedimento multivariado é sua efetiva falta de sensibilidade, quando comparado a técnicas univariadas. Enquanto os dois procedimentos uni e multivariado são assintoticamente equivalentes, para os tamanhos de amostra, encontrados na prática, a análise univariada convencional é mais poderosa que a análise multivariada, desde que estejam satisfeitas as suposições sobre Σ, pois as estimativas da análise univariada baseiam-se em um número maior de GL do que a análise multivariada correspondente. Mais ainda, o procedimento multivariado só pode ser aplicado nos experimentos em que o número total de unidades experimentais menos o número de tratamentos é maior ou igual ao número de medidas repetidas menos um, isto é, se $N-i \geq p-1$. Esta condição se faz necessária para que a matriz de somas de quadrados de produtos cruzados do erro (ε_{ijk}), associada aos efeitos dentro de sujeitos, seja não-singular. Nestas condições, esta matriz admite inversa, a qual é necessária para os cálculos das estatísticas multivariadas (Singer, Rocha e Nobre 2004).

O modelo estatístico utilizado, segundo Singer, Rocha e Nobre (2004), na análise multivariada de perfis talvez seja o mais simples dentre aqueles, usualmente, adotados para análise de medidas repetidas. Neste modelo as observações são expressas na forma matricial.

$$\underset{(nxp)}{y} = \underbrace{\begin{bmatrix} y_{111} & y_{112} & \cdots & y_{11p} \\ y_{121} & y_{122} & \cdots & y_{12p} \\ \vdots & \vdots & \vdots & \vdots \\ y_{1n_1 1} & y_{1n_1 2} & \cdots & y_{1n_1 p} \\ \vdots & \vdots & \vdots & \vdots \\ y_{g11} & y_{g12} & \cdots & y_{g_1 p} \\ \vdots & \vdots & \vdots & \vdots \\ y_{yn_g 1} & y_{gn_g 2} & \cdots & y_{gn_g p} \end{bmatrix}}_{\text{matriz dos dados}} = \underbrace{\begin{bmatrix} y'_{11} \\ y'_{12} \\ \vdots \\ y'_{1n_1} \\ \vdots \\ y'_{g_1} \\ \vdots \\ y'_{gn_g} \end{bmatrix}}_{\substack{\text{perfil de resposta da} \\ \text{matriz experiment al (ij)}}}$$

e modelo linear pode ser representado por:

$$\underset{(nxp)}{\varepsilon(y)} = \underset{(nxg)}{x} \, . \, \underset{(gxp)}{\beta} \, , \tag{35}$$

onde

$$\underset{(nxg)}{x} = \begin{bmatrix} 1_{n_1} & 0 & \cdots & 0 \\ 0 & 1_{n_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & I_{n_g} \end{bmatrix}$$

é uma matriz de especificação (planejamento) e

$$\underset{(gxP)}{\beta} = \begin{bmatrix} \mu_{11} & \mu_{12} & \cdots & \mu_{1p} \\ \mu_{21} & \mu_{22} & \cdots & \mu_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ \mu_{g1} & \mu_{g2} & \cdots & \mu_{gp} \end{bmatrix} = \begin{bmatrix} \mu'_1 \\ \mu'_2 \\ \vdots \\ \mu'_g \end{bmatrix}$$

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

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

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{12} & \sigma_2^2 & \cdots & \sigma_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_{1p} & \sigma_{2p} & \cdots & \sigma_p^2 \end{bmatrix}$$

onde σ_{ij} é a covariância entre as observações medidas em condições de avaliação distintas e σ_j^2 é a variação das observações medidas na mesma condição de avaliação (Singer, Rocha e Nobre 2004).

A fim de facilitar a interpretação da análise do modelo da equação (35) pode-se considerar a Figura 1 que corresponde à representação gráfica dos perfis médios observados de resposta em um caso particular com $g = 3$ tratamentos e $p = 4$ condições de avaliação.

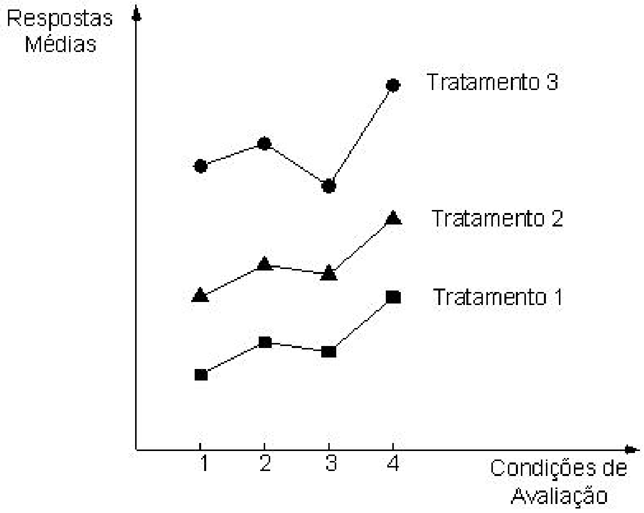

Figura 1 – Representação esquemática genérica entre tratamentos e condições de avaliação sem interação.

Conforme Andrade e Singer (1986), relativamente à configuração da Figura 1, os objetivos da análise de medidas repetidas podem ser traduzidas através das seguintes hipóteses:

H_{01}: os perfis médios de resposta correspondente aos diversos tratamentos são paralelos; não existe interação tratamento x condições de avaliação;

H_{02}: os perfis médios de resposta correspondente aos diversos tratamentos são coincidentes; não existe efeito de tratamentos;

H_{03}: os perfis médios de resposta correspondente aos diversos tratamentos são paralelos ao eixo da abscissa, isto é, não existe efeito das condições de avaliação.

Em termos dos parâmetros do modelo da equação (35) as hipóteses podem ser expressas por Perfis Paralelos:

$$H_{01} = \begin{bmatrix} \mu_{11} - \mu_{12} \\ \mu_{12} - \mu_{13} \\ \vdots \\ \mu_{1(p-1)} - \mu_{1p} \end{bmatrix} = \begin{bmatrix} \mu_{21} - \mu_{22} \\ \mu_{22} - \mu_{23} \\ \vdots \\ \mu_{2(p-1)} - \mu_{2p} \end{bmatrix} = \cdots = \begin{bmatrix} \mu_{g1} - \mu_{g2} \\ \mu_{g2} - \mu_{g3} \\ \vdots \\ \mu_{g(p-1)} - \mu_{gp} \end{bmatrix};$$

Perfis Coincidentes:

$$H_{02} = \begin{bmatrix} \mu_{11} \\ \mu_{12} \\ \vdots \\ \mu_{1p} \end{bmatrix} = \begin{bmatrix} \mu_{21} \\ \mu_{22} \\ \vdots \\ \mu_{2p} \end{bmatrix} = \cdots = \begin{bmatrix} \mu_{g1} \\ \mu_{g2} \\ \vdots \\ \mu_{gp} \end{bmatrix};$$

e Perfis Horizontais:

$$H_{03} = \begin{bmatrix} \mu_{11} \\ \mu_{21} \\ \vdots \\ \mu_{g1} \end{bmatrix} = \begin{bmatrix} \mu_{12} \\ \mu_{22} \\ \vdots \\ \mu_{g2} \end{bmatrix} = \cdots = \begin{bmatrix} \mu_{1p} \\ \mu_{2p} \\ \vdots \\ \mu_{gp} \end{bmatrix}.$$

Na forma da hipótese linear geral:

$$H = C.\beta.U = 0 \qquad (36)$$

onde C é uma matriz de constantes conhecidas com posto c, através da qual pode-se testar contrastes de tratamentos e U é uma matriz de constantes conhecidas com posto u, através da qual pode-se testar contrastes entre condições de avaliação.

Então tem-se as possíveis correspondências:

$$H_{01}:\ C_1 = \begin{bmatrix} 1 & -1 & 0 & \cdots & 0 \\ 1 & 0 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & \cdots & -1 \end{bmatrix} \text{ e } U_1 = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ -1 & 1 & \cdots & 0 \\ 0 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 \\ 0 & 0 & \cdots & -1 \end{bmatrix}$$

$H_{02} : C_1$ e $U_2 = I_p$

$H_{03} : C_3 = I_g$ e $U_3 = U_1$

3 RESULTADOS E DISCUSSÕES

Os resultados são apresentados seguindo-se a formulação proposta nos objetivos da pesquisa. Contudo, previamente, são discutidos alguns dados referentes à composição dos grupos das disciplinas. Assim, procede-se a identificação dos grupos, seguida das observações determinantes do desempenho, e, por último, realiza-se uma análise das principais constatações.

Da coleta de dados no levantamento feito no Colégio Técnico Industrial de Santa Maria, verificou-se que 62,3% dos alunos eram do gênero feminino, a grande maioria, 70,6%, possuía 14 anos, a maior parte deles (57,3%) são oriundos de escolas de Santa Maria e sendo que o número de alunos que provém de escola particular e da escola estadual praticamente é o mesmo.

Tabela 3 – Número e percentagem de alunos por gênero, idade, escola e cidade de origem.

Variáveis / Classificação	Números de alunos	Porcentagem (%)
Gênero		
Masculino	23	37,7
Feminino	38	62,3
Idade		
13 anos	9	14,7
14 anos	43	70,6
15 anos ou mais	9	17,7
Escola de origem ensino fundamental		
Estadual	25	41,0
Municipal	9	14,7
Federal	1	1,7
Particular	26	42,6
Cidade de origem		
Santa Maria	35	57,3
Outras localidades	26	42,7

Uma das perguntas na escola é sobre o desempenho do alunos nas dez disciplinas da primeira série do Ensino Médio oriundos de escolas pública ou particular e também referente ao gênero feminino e masculino.

Antes de aplicarmos uma análise de variância multivariada, tem-se que verificar os pressupostos de normalidade e de igualdade de matrizes de variância e covariância, para após se utilizar os dados. Para verificar a normalidade foi utilizado o gráfico Q-Q, descrição da página 23, que mostrou que não há grandes afastamentos dos pontos amostrais em relação à linha normal. O pressuposto de igualdade de matrizes de variância e covariância foi verificado pelo teste M de Box, conforme equação 22, e utilizou-se a aproximação pela distribuição F, equação (28), por ser a mais adequada, pois tem-se 2 grupos com tamanhos superiores a 20, e também o número de variáveis maior que 6, pois tinha-se 10 disciplinas.

Para verificar se existe diferença entre alunos oriundos de escola pública ou privada, realizou-se a análise de variância para dados repetidos ao nível de significância de 5% com os tipos de escolas (particular ou pública), conforme Tabela de dados no Anexo A, pode-se verificar (Tabela 4) que não há diferença significativa entre os tipos de escola, mas houve diferença significante entre as disciplinas. Não havendo interação significativa, verifica-se qual ou quais as disciplinas estão se diferenciando das demais.

Tabela 4 – Análise de variância para dados repetidos – variáveis escola e disciplina.

Efeito	SQ	GL	QM	F	p
Escola	0,12	1	0,12	0,019	0,890214
Erro	353,16	59	5,99		
Disciplina	287,53	9	31,29	80,530	0,000000
Disciplina x Escola	2,24	9	0,25	0,639	0,763640
Erro	206,35	531	0,39		

Através da Figura 2, a seguir, observa-se que as disciplinas se separam em grupos, onde a proximidade se dá pela semelhança de média, observa-se que as disciplinas FI, M e Q formam um grupo (Grupo 1), sendo que Química não difere significativamente de Física (Tabela 6), mas difere de Matemática, mas por ser sua média mais próxima de Matemática, decidiu-se deixar neste grupo por ser uma disciplina da área de exatas, que possuem uma média em torno de 7,11; As disciplinas B, LI e LB formam outro grupo (Grupo 2) onde a média está em torno de 7,99; e formando um último grupo (Grupo 3) as disciplinas EA, P, EF e LP com média em torno de 8,75.

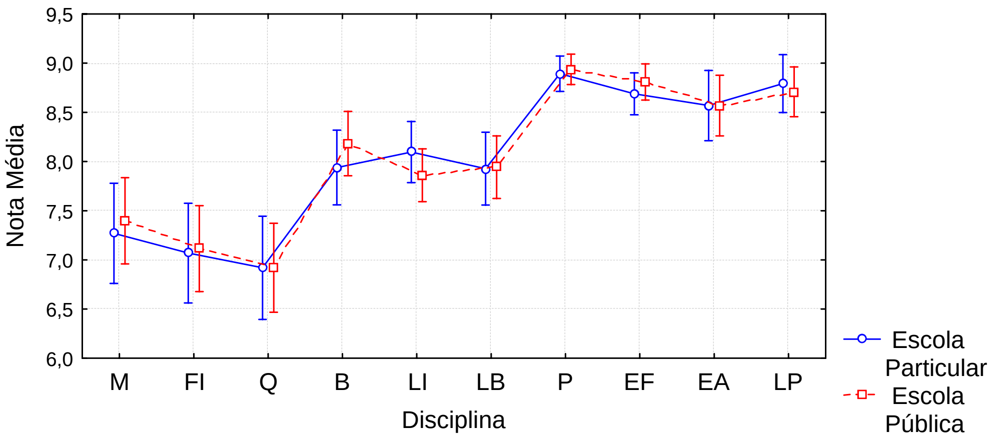

Figura 2 – Análise de perfis – variáveis escola e disciplina.

Após esta primeira verificação, onde se constatou que a origem escolar, (pública, particular) não influencia no desempenho dos alunos, e observando-se que havia grupos de disciplinas onde o desempenho poderia trazer uma diferenciação, verifica-se desta forma, na Tabela 5, as médias de cada uma das dez disciplinas, com seus respectivos desvios padrões.

Tabela 5 – Média e desvio padrão para as disciplinas.

Disciplina	Média	Desvio padrão
P	8,914725	0,400063
LP	8,750440	0,780034
EF	8,748516	0,563796
EA	8,568901	0,748242
B	8,060659	0,913779
LI	7,978077	0,717922
LB	7,934890	0,860672
M	7,333187	1,151594
FI	7,091758	1,295954
Q	6,919615	1,350555

Outra forma de se constatar as afirmativas anteriores é através da realização de teste de Tukey (Tabela 6) onde se pode concluir que não houve diferença entre as disciplinas do Grupo 1 (M, Q e FI), observando-se que Química difere de Matemática, mas por estar próximo da média, deixou-se num único grupo, Grupo 2 (B, LI, LB) e Grupo 3 (P, LP, EA e EF).

Tabela 6 – Teste de Tukey

Disciplina	Média				
P	8,914725	A*			
LP	8,750440	A			
EF	8,748516	A			
EA	8,568901	A			
B	8,060659		B		
LI	7,978077		B		
LB	7,934890		B		
M	7,333187			C	
FI	7,091758			C	D
Q	6,919615				D

* Disciplinas ligadas pela mesma letra não diferem significativamente pelo teste de Tukey, ao nível de 5% de probabilidade de erro.

Analisa-se, agora, se existe diferença significativa nas médias das disciplinas, quanto ao gênero, masculino ou feminino, conforme Tabela de dados no Anexo A.

Tabela 7 – Análise de variância para dados repetidos – variáveis disciplina e gênero.

Efeito	SQ	GL	QM	F	p
Gênero	28,95	1	28,95	5,267	0,025302
Erro	324,32	59	5,50		
Disciplina	263,34	9	28,67	74,820	0,000000
Disciplina x Gênero	5,14	9	0,57	1,491	0,147721
Erro	203,44	531	0,38		

Verifica-se na Tabela 7 que não existe interação entre disciplina e gênero, ou seja, o gênero que possui média maior o será em todas as disciplinas. Percebe-se também, neste quadro que existe diferença significante entre os gêneros, e que como já se sabe, entre as disciplinas.

Para sabermos qual dos gêneros é o que possui as melhores médias em todas as disciplinas, construiu-se o gráfico de perfis de médias que está representado na Figura 3.

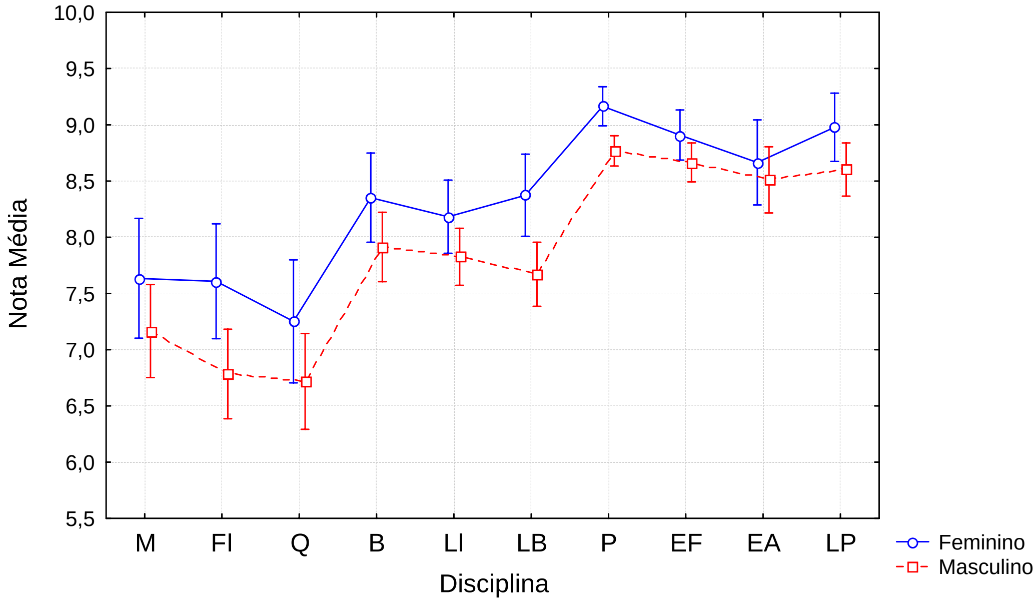

Figura 3 – Análise de perfis – variáveis disciplina e gênero.

Observa-se na Figura 3 que o gênero que possui melhor média em todas as disciplinas é o feminino. Além disso, pode-se observar que o comportamento entre os gêneros é o mesmo, ou seja, as médias das meninas, tanto quanto a dos meninos, são menores nas disciplinas de M, FI e Q.

4 CONCLUSÃO

A partir do presente estudo de caso e ao final das análises, observados os dados, verifica-se que não há diferenciação entre os tipos de escola (particular, pública), mas o fator de relevância passa a ser então as disciplinas que foram divididas em grupos e entre estes, constatou-se que há diferenciação.

As disciplinas separadas em grupos, onde a proximidade se dá pela semelhança das médias, observa-se que as disciplinas de Física, Matemática e Química formam um grupo, ou seja, as ciências exatas, que possuem uma média em torno de 7,11; Já as disciplinas de Biologia, Língua Inglesa e Literatura Brasileira formam outro grupo onde a média está em torno de 7,99; e em um terceiro grupo tem-se as disciplinas de Educação Artística, Projetos, Educação Física e Língua Portuguesa com média em torno de 8,75.

Observando-se o gênero, pôde-se constatar uma outra diferença, as alunas tiveram um desempenho superior ao dos alunos, em todas as disciplinas.

Por isso, conforme os resultados deste trabalho, apesar dos alunos serem oriundos de escola pública e escola particular, não se constatou diferença significativa no seu desempenho no primeiro ano do Ensino Médio do Colégio Técnico Industrial de Santa Maria, pois a diferença se torna visível nos grupos de disciplinas e no gênero dos alunos.

Como sugestão para trabalhos futuros, recomenda-se o acompanhamento, desses alunos na segunda e terceira séries do Ensino Médio e que a escola elabore e promova programas que miniminizem tais diferenças.

REFERÊNCIAS BIBLIOGRÁFICAS

ANDRADE, D.F. de.; SINGER, J. da M. **Análise de dados longitudinais.** (VII Simpósio Nacional de Probabilidade e estatística). 1ª ed. Campinas: Ed da USP, 1986. 106p.

HAIR, J. F.; et al. **Multivariate data analysis.** 5th.ed. New Jersey: Prentice-Hall, 1998.700p.

JOHNSON, R.A.; WICHERN, D.N. **Applied Multivariate Statistical Analysis.** 4th ed. New Jersey: Prentice-Hall, 1998. 686p.

LIMA, C.G. de. **Análise de dados longitudinais:** provenientes de experimentos em blocos casualizados. SP, 1996. 126f. Tese (Doutorado em Agronomia) – Universidade de são Paulo, são Paulo, 1996.

MARDIA, K.V.; KENT, J.T.; BIBBY,J.M. **Multivariate Analysis.** San Diego: Academic Press, Inc, 1994.518p.

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

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

RIBOLDI, J.; et al. **Modelos Mistos e Medidas Repetidas.** Universidade Federal do Rio Grande do Sul. Instituto de Matemática. Departamento de Estatística. Caderno de Matemática e Estatística, 2002, não-paginado, impresso.

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

SINGER, J. da M., ROCHA, F. M. da, NOBRE, J. S. **Análise de medidas repetidas.** IV Jornada Regional de Estatística. Universidade Estadual de Maringá, Maringá, 2004.

STUKER, H. **Análise multivariada para dados onde a característica observada é subdividida em K classes.** SP, 1986. 92f. Dissertação (Mestrado em Agronomia) – Universidade de São Paulo, São Paulo, 1986.

VIEIRA, S. **Estatística Experimental.** São Paulo: Atlas, 2 ed., 1999, 185p.

ANEXO

Anexo A – Tabela de dados, gênero feminino e masculino, escola particular e pública.

Gênero	Escola	M	FI	Q	B	LI	LB	P	EF	EA	LP
F	Particular	9,3	9	9,3	9,2	9,1	9,4	9,7	8,7	9,5	9,6
F	Particular	7,2	6,3	7,2	8,7	8,3	8,9	9	9	9	9
F	Particular	9,3	9,7	9,2	9,6	9,6	9,4	10	9,5	9,8	9,8
F	Particular	6,5	6,5	5,6	7,6	7,6	8,4	9	9,1	9,3	7,6
F	Particular	7,3	6,4	6,3	7,4	7,3	8,2	8,9	8,6	9	9,1
F	Particular	6,2	5,6	5,5	7,8	7,7	8,7	8,8	8,6	8,9	8,3
F	Particular	7,2	8,3	6,8	7,2	8,1	7,4	8,9	8,8	8,2	8,5
F	Particular	6,5	7,8	6,8	8,7	8,9	8,7	9	8,8	9,2	7,9
F	Particular	8,5	8,4	7,9	8,8	8,7	7,8	9,4	9,2	8,2	9,4
F	Particular	6,6	7,1	6,7	6,9	7,3	7,7	9,1	8,9	7,6	8,4
F	Particular	7,7	8,2	7,5	6,9	8,1	7,8	9,3	8,9	7,4	9,4
F	Pública	6,7	5,5	5,9	8,2	7,3	8,1	8,6	8,3	9,1	8,1
F	Pública	7,4	7	7,6	8,8	8,6	8,7	9	9,4	8,5	9,2
F	Pública	6,8	6,4	7,1	8,1	8,1	7,8	9	8,4	8,3	9
F	Pública	8,5	8,1	8,4	9,7	9,2	9,7	9,6	9,3	9,6	9,7
F	Pública	7,8	7,2	7,1	9	8	8,3	9,3	9	8	8,6
F	Pública	9,1	8,3	8	8,2	7,9	8	9,1	8,4	7,9	9,7
F	Pública	8,6	9,1	8,1	8,7	8,9	8,5	9,3	8,7	8,9	9,4
F	Pública	7,9	8,2	7,2	8,7	7,9	8	9,2	9,2	8,8	9
F	Pública	7,9	8,4	7,4	8,9	8	8,8	9,4	9	7,8	9,4
F	Pública	6,4	7,1	5,3	8	7	7,3	8,7	8,5	7,6	9,3
F	Pública	8,2	9,1	8,2	8,5	8,4	8,4	9,4	9,2	9,5	9,6
F	Pública	8	7,3	7,7	8,5	8,2	8,6	9,1	9,4	9,2	8,5
M	Particular	8,2	6,3	6,4	8,6	8,3	8,5	8,7	9	9,4	8,5
M	Particular	8,6	7	7,9	8,6	9	8,4	8,9	9,2	9,8	9,6
M	Particular	5,5	5,8	6,6	8	8,1	7,4	9,1	8,9	8,8	8,7
M	Particular	7,1	6,5	7,3	8,4	8,6	8,4	9,3	8,4	9,5	9,5
M	Particular	3,1	3,8	4	5,9	5	4	7,4	7,1	5,1	7,4
M	Particular	6,6	6,6	5,3	7,4	7,5	7,4	8,3	7,5	7,2	7,7
M	Particular	8,5	7	7,5	7,7	7,3	7,3	8,7	8,9	9	9,3
M	Particular	5,5	5,8	5,7	7,5	7,6	7,1	8,6	8,6	7,5	8
M	Particular	7,3	6,2	7,2	9	8,5	8,4	8,6	9	9,6	8,1

Tabela de dados, gênero feminino e masculino, escola particular e pública (continuação).

Gênero	Escola	M	FI	Q	B	LI	LB	P	EF	EA	LP
M	Particular	5,6	7,5	5,7	7	8,3	7,5	8,6	8,9	9	8,5
M	Particular	8,4	8,5	7,5	7,6	7,9	7,9	9	8,3	8	9,3
M	Particular	5,9	6	5	5,4	7,9	6,7	7,8	8	7	8,9
M	Particular	8,9	8,6	9	8,4	8,2	8,1	8,9	8,5	8,6	9,4
M	Particular	8,9	7,6	7,5	8,6	8,4	8,2	8,8	8,9	8,6	9,2
M	Pública	9,5	9	9,4	9,1	8,9	9,2	9,3	9,4	9,8	9,7
M	Pública	6,7	5,4	5,5	7,8	7,8	7,5	9	9	8,3	8,8
M	Pública	9,2	8,4	8,8	9,5	9,2	9	9,4	9,3	9,8	9,3
M	Pública	5,2	5	6,3	7,1	7,1	7,1	9	8,6	8,1	8,8
M	Pública	7,4	7	7,4	8,4	7,2	7,1	7,9	8,4	7,3	8,8
M	Pública	7,2	6,5	7,1	8,1	7,1	7,7	8,5	9	8,9	8,8
M	Pública	9	8,3	8,6	8,8	8,3	8,2	9,1	9,2	8,9	8,6
M	Pública	8,3	7,8	7,9	8,5	8,7	8,8	9,2	8,7	9,3	8,8
M	Pública	5,5	6,1	4,8	8,8	7	6,6	8,2	7,2	7,2	7
M	Pública	5,9	5,2	5,8	6,6	7,1	7	8,7	8,7	8,2	8,6
M	Pública	7	6,5	5,7	8,4	7,1	8,7	9	8,8	8,7	8,1
M	Pública	6,4	4,9	5,7	7,6	7	7,4	8,9	8,5	8,3	8,3
M	Pública	7,8	7,4	7,8	9,3	8,1	9	9,5	9,3	9,5	9,6
M	Pública	8,6	8,2	8,1	8,5	9	8,3	9	9,8	9,2	9,2
M	Pública	6,8	5	5,5	7,4	7,1	8,1	8,5	9,4	7,9	7,1
M	Pública	5,7	7	5,6	7,2	7,5	7,1	8,6	8,9	8,7	7,3
M	Pública	7,1	7	5,5	8,8	7	7,9	8,8	8,7	9,1	8,7
M	Pública	7,3	7,4	7,5	8,2	7,7	7,6	9,1	8,8	9,2	8,9
M	Pública	6	6,7	3	5,2	7,6	5,6	8,2	7	7,8	7
M	Pública	8,1	7,8	7,9	7,1	8,4	8	8,6	8,4	7	8,8
M	Pública	5,3	4,5	6,1	6,6	7,3	6,5	8,4	8,9	8,6	7,4
M	Pública	8,3	8,7	7,5	8,3	8,4	8,3	9,2	8,8	9	9,4
M	Pública	7,3	7,5	6,7	7,8	7	7,1	9	8,7	7,9	8,3