Universidade Federal de Minas Gerais
Instituto de Ciências Exatas
Departamento de Estatística

Marcos Paulo Patussi dos Santos

# Análise dos desempenhos obtidos pelos alunos do Ensino Médio Da Fundação de Ensino de Contagem, avaliando as disciplinas de Língua Portuguesa e Matemática.

Belo Horizonte
2011

Marcos Paulo Patussi dos Santos

# Análise dos desempenhos obtidos pelos alunos do Ensino Médio Da Fundação de Ensino de Contagem, avaliando as disciplinas de Língua Portuguesa e Matemática.

Trabalho apresentado ao Curso de Estatística do ICEX-UFMG para obtenção do grau de Especialista em Estatística.

Aluno: Marcos Paulo Patussi dos Santos
Orientador: Prof. Ela Mercedes M de Toscano

Belo Horizonte
2011

# Análise dos desempenhos obtidos pelos alunos do Ensino Médio Da Fundação de Ensino de Contagem, avaliando as disciplinas de Língua Portuguesa e Matemática.

**Marcos Paulo Patussi dos Santos**

**Prof. Ela Mercedes M de Toscano**

Departamento de Estatística da Universidade Federal de Minas Gerais

**Outubro, 2011**

## I. RESUMO

A FUNEC (Fundação de Ensino de Contagem) é uma autarquia municipal, onde a prefeitura atende o nível do ensino médio regular e técnico, estaremos fazendo uma análise dos desempenhos obtidos pelos alunos do Ensino Médio da Fundação de Ensino de Contagem, avaliando as disciplinas de Língua Portuguesa e Matemática, observando o rendimento dos alunos nos três últimos anos do ensino médio.

Esse estudo surgiu na necessidade de avaliar a real condição que os alunos do ensino médio chegam à universidade, bem como verificar o comportamento do rendimento das disciplinas de Língua Portuguesa e Matemática e uma provável relação existente entre essas duas disciplinas. Por estar diretamente ligado a sala de aula observei durante esses dez anos de docência o desenvolvimento e comportamento dos alunos em relação ao seu desempenho no conteúdo de matemática. No ano de 2009 e 2010 venho desenvolvendo junto ao corpo de professores da escola na qual eu trabalho um projeto de intervenção pedagógica na leitura, visando melhorar a interpretação desses alunos dentro do conteúdo específico, no caso a matemática, com o objetivo de proporcionar uma melhora no rendimento e desempenho escolar, uma vez que as análises iniciais demonstram uma necessidade de interpretação da escrita para ai sim aplicarmos as ferramentas da matemática, espera que possa demonstrar essa real necessidade que presencio em sala de aula.

Palavras-chave:(Desempenhos,analise, Matemática, língua Portuguesa)

Análise dos desempenhos obtidos pelos alunos do Ensino Médio Da Fundação de Ensino de Contagem, avaliando as disciplinas de Língua Portuguesa e Matemática.

**Marcos Paulo Patussi dos Santos**

**Prof. Ela Mercedes M de Toscano**

Departamento de Estatística da Universidade Federal de Minas Gerais

**Outubro, 2011**

## Abstract

The FUNEC (Count Education Foundation) is a local authority, where the city meets the level of regular and high school coach, we will be doing an analysis of the performance achieved by students in the School of Education Foundation Counting, weighing the subjects Portuguese and Mathematics, watching the students' performance in the last three years of high school.
This study arose the need to assess the real condition that high school students reach university, and to verify the behavior of the performance of the subjects of Portuguese Language and Mathematics and a probable relationship between these two disciplines. Since this is directly connected to the classroom I observed during these ten years of teaching development and student behavior in relation to their performance in mathematics content. In 2009 and 2010 have been developing along the body of school teachers that I work in a pedagogical intervention project in reading, to improve the interpretation of these students into content-specific in mathematics, in order to provide an improvement income and school performance, since initial analysis showed a need for interpretation of writing for oh yes we apply the tools of mathematics, I hope you can demonstrate that real need that I witness in the classroom.

Keywords: (Performances , analysis, Mathematics, Portuguese language)

Análise dos desempenhos obtidos pelos alunos do Ensino Médio Da Fundação de Ensino de Contagem, avaliando as disciplinas de Língua Portuguesa e Matemática.

## I - INTRODUÇÃO

Estes estudos têm como objetivo mostrar os desempenhos dos alunos da Fundação de Ensino de Contagem (FUNEC), em matemática e em língua Portuguesa, bem como o desempenho por gênero em cada disciplina e uma possível relação entre o desempenho das duas disciplinas, esse trabalho terá como foco aplicação de ferramentas estatísticas, para apresentação de resultados e análises de variáveis. Espera-se ainda avaliar o desenvolvimento desses alunos no transcorrer dos três anos do ensino médio, mostrando o desempenho e aproveitamento, visto que o desempenho destes alunos no decorrer dos últimos dez anos pela prática de sala de aula vem mostrando uma piora significativa no transcorrer destes anos, tanto no aspecto de bagagem de conteúdo como no aspecto disciplinar, que acaba puxando cada dia mais os índices dos resultados para baixo. Percebe-se ainda que o desempenho dos alunos do sexo feminino vêem apresentando um desempenho melhor que o sexo masculino, e a grande necessidade de se trabalhar a interpretação da língua portuguesa para a aplicação dentro do conteúdo da matemática. Espera-se demonstrar com a análise desses resultados estas relações dentro do ensino médio.

# II - METODOLOGIA

Este trabalho está dividido em três fases: Na primeira fase, vamos descrever o rendimento em matemática e português, ano a ano, usando estatísticas descritivas, analisando média, coeficiente de variação, quartis e outras estatísticas relevantes, no segundo momento vamos descrever a relação entre rendimento em matemática e Língua Portuguesa, usando estatísticas para dados bivariados como, por exemplo, correlação, e no terceiro momento iremos aplicar o teste qui quadrado e índices para verificarmos a dependência e a comparação do rendimento dos alunos ano a ano.

## 2.1 - Estatísticas Descritivas

### 2.1.1 Medidas de tendência central;

A mediana é o valor médio central depois de ordenados os dados em forma ascendente. As propriedades da mediana podem ser resumidas como segue:

· A mediana não é particularmente sensível a valores extremos.

· A mediana poderá tomar sempre um único valor.

· A mediana será igual a um valor observado se o número de observações n é um número impar.

· Para calcular a mediana, não usamos todas as observações.

**2.1.2** A **média aritmética**, o conceito da média aritmética ou simplesmente média, é bastante familiar. Para calcular a média, se soma todas as observações e divide pelo o número de valores somados. Matematicamente, se n observações é representado como: $X_1$, $X_2$, ... $X_n$, a média aritmética pode ser escrito como:

$$\overline{X} = \frac{X_1 + X_2 + \quad \ldots X_n}{n}$$

As propriedades da **média aritmética** podem ser resumidas como segue:

- · Para calcular a média se usa todas as observações disponíveis
- · A média é afetada por valores extremos.
- · A média é uma medida estável a pequenas mudanças das observações.
- · A média não necessariamente será igual a um dos valores observados.
- · A média não pode ser determinada graficamente.

**Medidas de Dispersão**

Outra medida de dispersão é definida com os desvios quadráticos (desvio)$^2$.

- **<u>2.1.3 Variância Amostral</u>** ($S^2$)

$$S^2 = \frac{\sum (X_i - \overline{X})^2}{N}$$

- **<u>2.1.4</u>** **<u>Desvio Padrão Amostral</u>** (s,)

$$S = \sqrt{S^2}$$

Para Entender o Desvio Padrão:

- Devemos ter em mente que o desvio padrão mede a variação entre valores.
- Valores próximos uns dos outros originam desvios padrão menores, enquanto valores muito afastados uns dos outros dão um desvio padrão maiores.
- Uma regra prática que utiliza a amplitude para obter uma estimativa bastante rudimentar do desvio padrão, é como segue:
- Uma regra prática: Desvio padrão ≈ $\frac{amplitude}{4}$

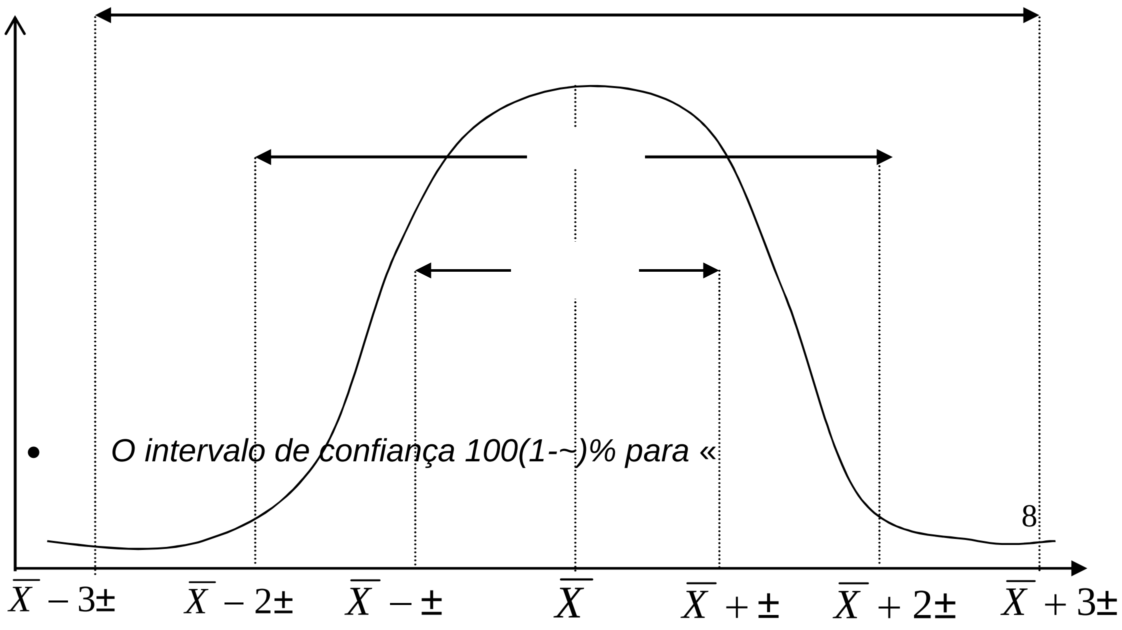
Desde que conhecemos o desvio padrão podemos utilizá-lo para entender melhor os dados, fazendo estimativas dos valores mínimos e máximos como segue:

Mínimo ≈ (média) – 2 x (desvio padrão)

Máximo ≈ média + 2 x (desvio padrão).

Outra regra que auxilia a interpretação do valor de um desvio padrão é a regra empírica, aplicável somente a conjuntos de dados com distribuição aproximadamente em forma de sino

Regra empírica, aplicável somente a conjuntos de dados com distribuição aproximadamente simétrica (forma de sino).




- *O intervalo de confiança 100(1-~)% para «*

$$\left[ \overline{X} - t_{n-1}\frac{\pm}{\sqrt{n}} \; ; \; \overline{X} + t_{n-1}\frac{\pm}{\sqrt{n}} \right]$$

*Onde $t_{n-1}$ é o percentil (1 - ~/2) da distribuição de Student com n -1 graus de liberdade.*

- **<u>2.1.5 Coeficiente de Variação</u>** (CV)

$$CV = \left( \frac{S}{\overline{X}} \right) 100\%$$

**Onde:**

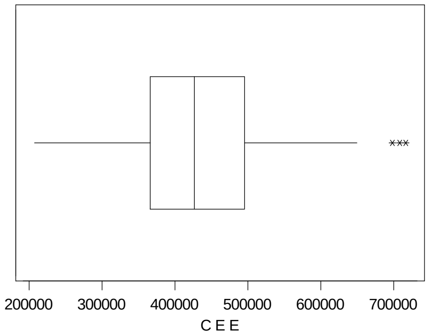
$$\overline{X} = \frac{\sum X}{n}$$ é a média aritmética e

$$S^2 = \frac{\sum (X_i - \overline{X})^2}{n-1}$$ é o desvio padrão amostral.

**Medidas Separatrizes**

- **Quartis**: Divide a distribuição em 4 partes iguais. Há três quarties denotados por $Q_1$, $Q_2$ e $Q_3$, que divide os dados ordenados em 4 grupos com 25% deles em cada grupo;
- **Decis**: Divide a distribuição em 10 partes iguais. Há nove Decies, denotados por $D_1$, $D_2$, ... $D_9$, que dividem os dados em 10 grupos com cerca de 10% deles em cada grupo.
- **Percentis**: Divide a distribuição em 100 partes iguais. Há 99 percenties, que dividem o dado em 100 grupos com cerca de 1% em cada grupo.

- **2.1.6 Diagrama em caixas – Box – Plot**




Os valores maiores do que $Q_3 + 1,5(Q_3 - Q_1)$ ou os valores menores do que $Q_1 - 1,5(Q_3 - Q_1)$ são considerados atípicos.
Os valores maiores do que $Q_3 + 3(Q_3 - Q_1)$ ou os valores menores do que $Q_1 - 3(Q_3 - Q_1)$ são considerados outliers.

- **2.1.7 Coeficiente de Correlação Linear**

Uma das estatísticas que quantifica o grau de relacionamento linear entre duas variáveis ou series é o coeficiente de correlação ($\rho$). O coeficiente de correlação amostral é dado por:

$$r = \rho(Y, X) = \frac{\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)\left(X_i - \overline{X}\right)}{\sqrt{\sum_{i=1}^{n}(X_i - \overline{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}}$$

onde $x_1, x_2, \ldots, x_n$ e $y_1, y_2, \ldots, y_n$ são os valores medidos de ambas as variáveis, e $\overline{X}$ e $\overline{Y}$ são as médias aritméticas de ambas as variáveis

Propriedades do coeficiente de correlação linear:

- O valor de ***r*** está sempre entre -1 e 1
- O valor não varia se todos os valores de qualquer uma das variáveis são convertidos por uma escala diferente.
- Permutando todos os valores de X e Y, ***r*** permanecerá inalterado.
- ***r*** não serve para medir a intensidade de um relacionamento não linear.

### 2.1.8 Testes de Hipóteses para a correlação linear

O teste formal de hipóteses é para determinar se existe correlação significativa entre duas variáveis ou series. A hipótese nula e alternativa se expressará como segue:

$$H_0 : \rho = 0$$
$$H_1 : \rho \neq 0$$

Estatística do teste

$$t = \frac{(r - \mu_r)}{S_r} \sim t_{(n-2, 1-\alpha/2)}$$

Onde a média e desvio padrão amostral de ***r*** é dado:

$$\mu_r = E(r) \qquad\qquad S_r = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}}$$

Como supomos que $\rho = 0$, decorre que. $\mu_r = 0$ Mostra-se também que o desvio padrão de ***r*** pode ser expresso como

$$S_r = \sqrt{\frac{(1-r^2)}{(n-2)}}$$

Podemos usar a seguinte estatística do teste

$$t = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}} \sim t_{(n-2, 1-\alpha/2)}$$

Rejeita-se a hipótese nula com nível $\alpha$ se $|t_c| > t_{n-1, 1-\alpha/2}$

### 2.1.9 Comparação de duas Médias

Amostras independentes com variâncias desconhecidos e iguais, $\sigma_1 = \sigma_2$ :

Seja X a variável aleatória que representa a característica de interesse em cada uma das populações. O tamanho das amostras aleatórias independentes $n_1$ e $n_2$ podem eventualmente ser iguais.

Estimador pontual de $\mu_1$ - $\mu_2$ é $\bar{x}_1 - \bar{x}_2$. Se ambas as populações tiverem uma distribuição normal, ou se os tamanhos de amostra forem suficientemente grades a ponto de o Teorema Central nos permitir concluir que

$$\bar{x}_1 - \bar{x}_2 \sim Normal\left( (\mu_1 - \mu_2), \sigma^2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right) \right)$$

Considerando que $S_1^2 \, e \, S_1^2$ são estimadores não viciados dessa variância, as usaremos para estimar uma combinação de variâncias:

$$S_C^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{(n_1 - 1) + (n_2 - 1)}$$

Assim, uma estimação por intervalo da diferença de médias assumirá a seguinte forma

$$\bar{x}_1 - \bar{x}_2 \pm t_{\alpha/2} S_c \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

onde (1 - $\alpha$) é o coeficiente de confiança.

Se nosso interesse é testar:

$H_0$ As médias populacionais são iguais;

$H_1$: As médias populacionais as médias não são iguais

As hipóteses em termos das medias populacionais $\mu_1$ e $\mu_2$;

$$H_0 : \mu_1 = \mu_2$$
$$H_1 : \mu_1 \neq \mu_2$$

A estatística de teste para testes de hipóteses sobre $\mu_1$ - $\mu_2$ com variâncias desconhecidos e $\sigma_1 = \sigma_2$ é a seguinte:

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{S_C \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{(n_1 + n_2 - 2)}$$

Estatística do teste *com variâncias desconhecidos,* $\sigma_1 \neq \sigma_2$ é a seguinte*:*

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \sim t_{(v)}$$

onde v é o menor valor entre ($n_1$-1) e ($n_2$-1); os graus de liberdade são corrigidos pela expressão:

$$v = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{\left(S_1^2/n_1\right)^2}{n_1 - 1} + \frac{\left(S_2^2/n_2\right)^2}{n_2 - 1}}$$

**2.2 Testes de Igualdade de Variâncias**

Sejam X e Y as variáveis aleatórias representando as características de interesse em cada uma das populações, tais que X ~ Normal ($\mu_x$,$\sigma_x^2$) e Y ~ Normal ($\mu_y$,$\sigma_y^2$). O tamanho das amostras aleatórias independentes $n_1$ e $n_2$ podem eventualmente ser iguais, se queremos testar a hipóteses:

$$H_0 : \pm_1^2 = \pm_2^2$$
$$H_1 : \pm_1^2 \neq \pm_2^2$$

a estatística de teste é

$$F = \frac{S_x^2}{S_y^2} \sim F_{[(n_1-1),(n_2-1)]}$$

sob a hipótese nula, pode ser mostrado que F segue uma distribuição de Fisher-Snedecor. Por tanto se $F_{ob}$ € RC, rejeitamos a hipótese de igualdade das variâncias.

**Diferenças entre Proporções**

Supondo que duas amostras independentes foram retiradas, uma de cada população, e seja $x_1$ o número de sucessos obtidos em uma amostra de $n_1$ observações e $x_2$ o número de sucessos em uma amostra de $n_2$ observações, obtemos as estimativas pontuais das proporções:

$$\hat{p}_1 = \frac{x_1}{n_1} \quad e \quad \hat{p}_2 = \frac{x_2}{n_2}$$

Um estimador pontual da diferença entre proporções amostrais é dado por. A distribuição desta diferença de proporções será aproximada pela Normal

$$\hat{p}_1 - \hat{p}_2 \sim Normal\left( (p_1 - p_2), \left( \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2} \right) \right)$$

Em forma geral uma estimativa por intervalo assumirá a seguinte forma

$$\hat{p}_1 - \hat{p}_2 \pm z_{\sim/2} \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

onde (1 - $\alpha$) é o coeficiente de confiança.

### 2.2.1 Teste de Hipóteses sobre Diferenças entre Proporções

Consideremos que as duas amostras foram retiradas, uma de cada uma das populações. O tamanho das amostras aleatórias independentes $n_1$ e $n_2$ podem eventualmente ser iguais, se queremos testar:

$H_0$ As proporções populacionais são iguais;

$H_1$: As proporções populacionais não são iguais

As hipóteses em termos das proporções populacionais $p_1$ e $p_2$

$$H_0 : p_1 = p_2$$
$$H_1 : p_1 \neq p_2$$

Sendo a hipótese nula verdadeira, as proporções populacionais são iguais, denotando seu valor comum por p, isto é, $p_1 = p_2 = p$, então podemos obter um estimador para p

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2}$$

Neste caso o estimador do desvio padrão amostral $\hat{p}_1 - \hat{p}_2$ torna-se:

$$\pm_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}} = \sqrt{p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

e a estatística de teste para testes de hipóteses para testes de hipóteses sobre $p_1 - p_2$:

$$z = \frac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

Essa estatística de teste aplica-se a situações com grandes amostras, em que

$$n_1 p_1,\, n_1(1-p_1),\, n_2 p_2 \;e\; n_2(1-p_2)$$

são todas maiores ou iguais a 5.

## III-ANÁLISE DO BANCO DE DADOS

Os dados analisados foram colhidos da Fundação de Ensino de Contagem, na unidade Hibisco, contendo turmas do 1º, 2º e 3º anos, têm como anos de referência (2003 a 2008) as escolhas das turmas foram aleatórias, a média de alunos por cada turma foi de 38 alunos. De cada aluno selecionado se observará a nota nas disciplinas de matemática e português, e sua condição ao final do ano, aprovado ou reprovado.

Na descrição e análise dos dados amostrais, usaremos estatística descritiva, diversos gráficos e histogramas para verificar as diversas hipóteses de comparação de médias, serão realizados testes de hipóteses. A descrição da associação e independência das variáveis, Também será utilizada tabelas de contingência e estatísticas do teste, para verificar a dependência e associação entre variáveis.

### 3.1-Descrição da amostra

. A partir dos dados amostrais, procurou-se observar a evolução do rendimento dos estudantes em matemática e português por ano e por série e sua situação final global por ano (aprovado, reprovado).

*3.2-Análise do rendimento em matemática*

Na tabela 01, encontra-se um resumo das estatísticas descritivas do rendimento em matemática no período em estudo, pode-se observar que a média no geral está muito baixa, puxada pelas notas zero dos alunos, só em 2004 e 2005 se tem notas médias acima de 60. A figura 01 mostra a evolução das notas médias no período, observa-se que as médias esta abaixo de 60, chegando a 28,48 em 2007, as notas mínimas de zero com exceção nos anos de 2004 e 2005, com notas máximas em torno de 90. A figura 02 mostra a grade variabilidade das notas, com dados discrepantes inferiores,

Tabela 01: Estatísticas descritivas das notas de matemática

| Variável | N | Média | Desv. Padrão | Coef. Var. | Mínimo | Máximo |
|---|---|---|---|---|---|---|
| Mat 03 | 48 | 54,96 | 31,84 | 58% | 0,0 | 88,00 |
| Mat 04 | 42 | **70,52** | 22,15 | 31% | 21,00 | **90,00** |
| Mat 05 | 38 | **71,39** | 22,45 | 31% | 21,00 | **90,00** |
| Mat 06 | 46 | 45,83 | 26,29 | 57% | 0,0 | 80,00 |
| Mat 07 | 40 | **28,48** | 26,14 | 92% | 0,0 | 84,00 |
| Mat 08 | 44 | 51,45 | 23,50 | 46% | 0,0 | 93,00 |

### 3.1.1 Comparação entre as médias de Matemática e Língua Portuguesa

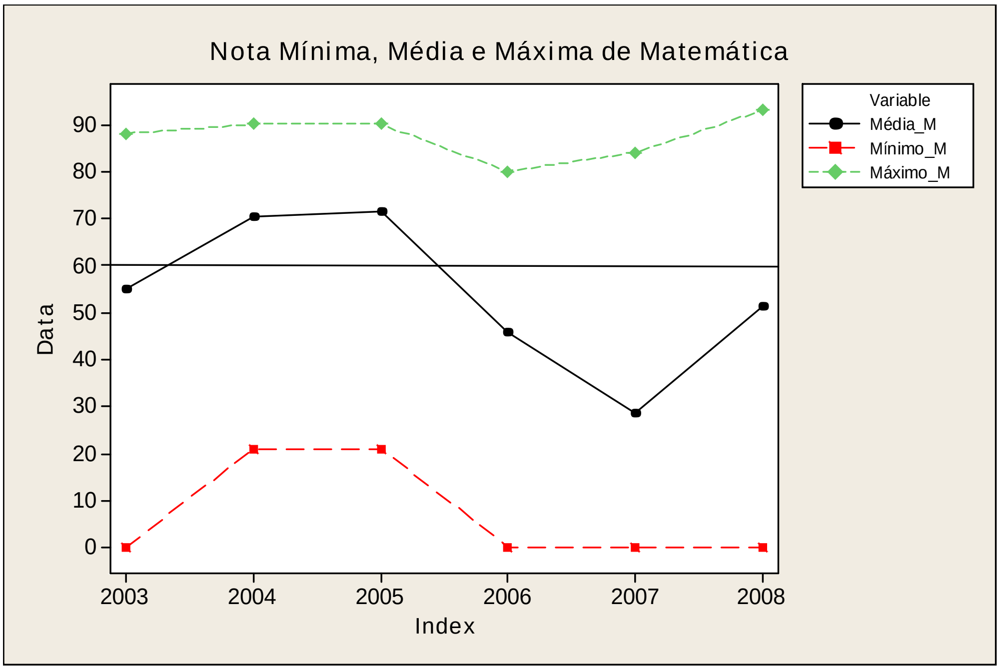



Figura 01: Notas de matemática no período de 2003 a 2008

A média do desempenho de matemática apresentou um resultado melhor nos anos de 2004 e 2005, verifica-se ainda que no ano de 2007 foi o pior ano e desempenho de matemática, pois as médias foram muito baixas.

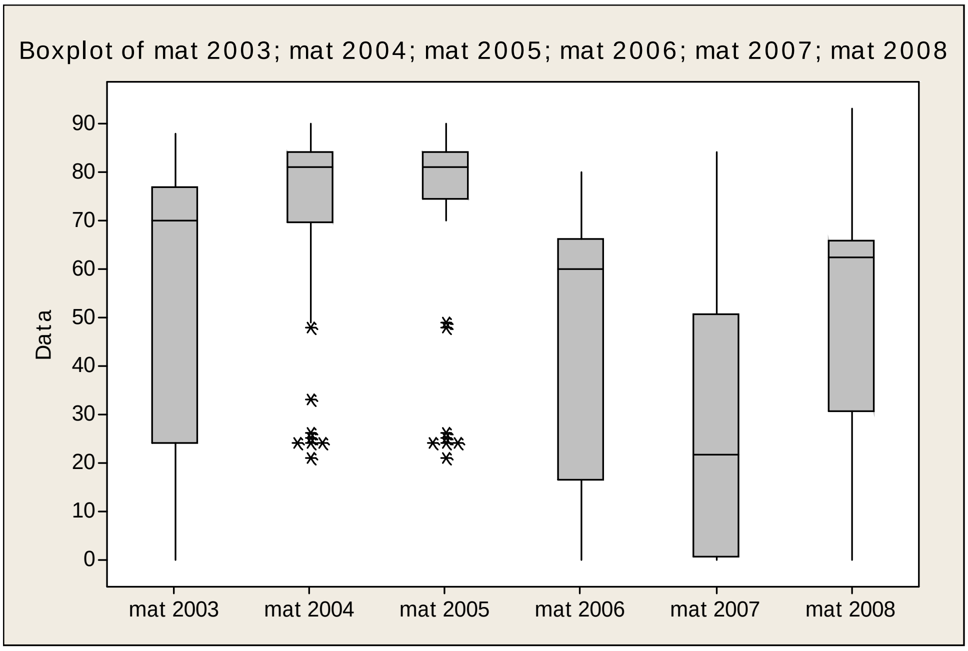



Figura 02: Notas de matemática no período de 2003 a 2008

## 4. Desempenho de Matemática

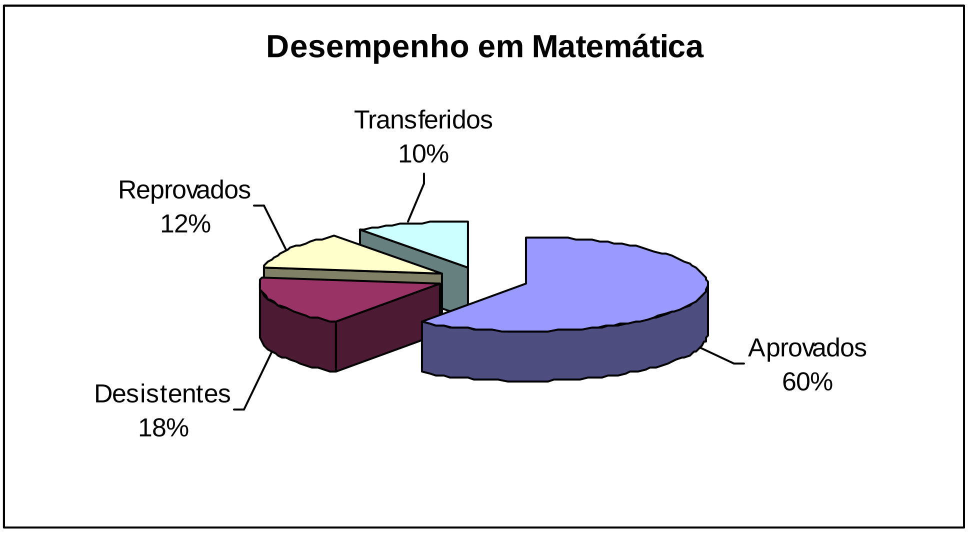



Figura 03: Situação final na disciplina matemática no período de 2003 a 2008

### 4.1 Agora veremos o aproveitamento, segundo o sexo:

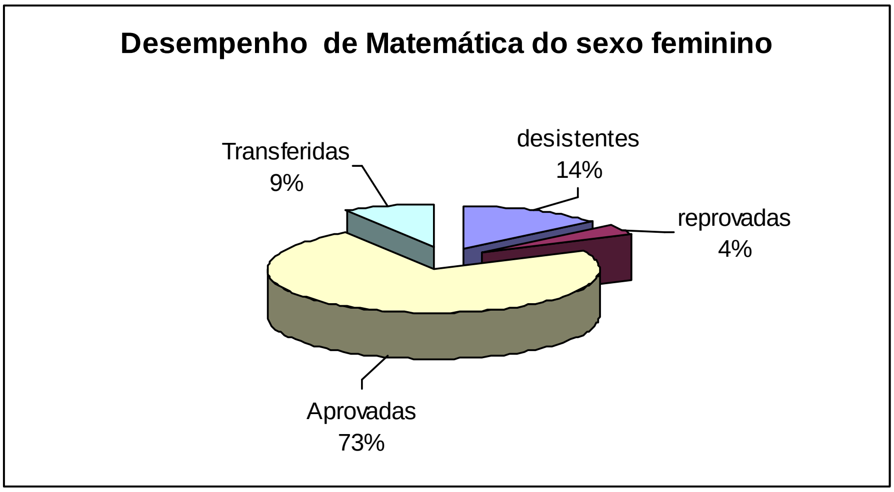



Figura 04: Período de 2003 a 2008

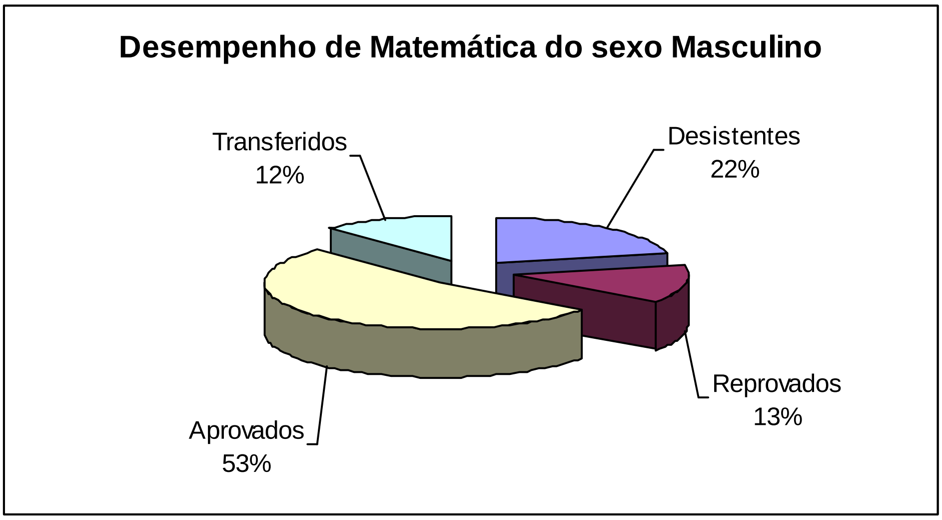



Figura 05: Período de 2003 a 2008

Comparando as figuras 03 e 04, podemos observar que o desempenho obtido para o sexo masculino é inferior ao sexo feminino em ambos os conteúdos, em matemática a diferença é de 20% no índice de aprovação, 9%

no índice de reprovação, 3% no índice de transferência e 8% no índice de desistentes, ou seja, valores bem significativos.

## 5. Proporção entre aprovados e reprovados do conteúdo de Matemática e Língua Portuguesa .

| PORT \ MAT | Reprovados | Aprovados | Total |
|---|---|---|---|
| Reprovados | 63 | 13 | 76 |
| Aprovados | 91 | 599 | 690 |
| Total | 154 | 612 | 766 |

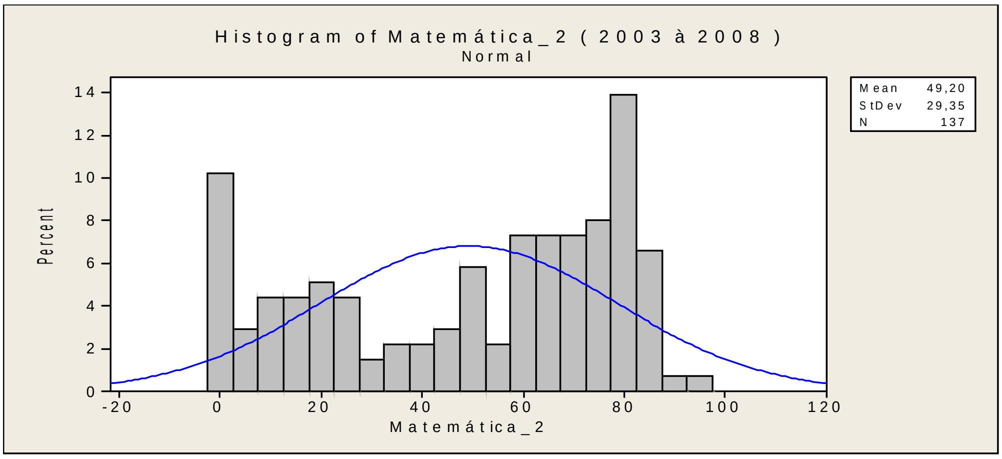



Figura 06: Distribuição do período de 2003 a 2008

Observando a distribuição acima, verificamos que as notas apresentam-se distribuídas em blocos, com algumas discrepâncias de valores abaixo da média, observa-se também um valor alto do desvio padrão que representa uma variabilidade bastante significativa.

## 6.-Análise do rendimento em português

Na tabela 02, encontra-se um resumo das estatísticas descritivas do rendimento em língua Portuguesa no período em estudo, pode-se observar que a média no geral está muito baixa, observa-se ainda quem em nenhum dos anos observados a média ultrapassou os 60% e aproveitamento, uma vez novamente puxada pelas notas zero dos alunos, que representam uma parcela significativa de abandonos, com valores bastante representativos como mostra a figura. A figura 06 mostra a evolução das notas médias no período, observa-se que as médias estão abaixo de 60, somente nos anos de 2003 a 2007 os resultados chegam próximos aos da média, observa-se ainda que nos anos de 2004 e 2005 não ocorreram notas 0, com notas máximas em torno de 90. A figura 07 mostra uma variabilidade das notas, com dados discrepantes inferiores,

Tabela 03: Estatísticas descritivas das notas de português

| Variável | N | Média | Desv. Padrão | Coef. Var. | Mínimo | Máximo |
|---|---|---|---|---|---|---|
| Port03 | 48 | 55,15 | 32,47 | 59% | 0,0 | 89,00 |
| Port04 | 42 | 58,74 | 23,09 | 39% | 15,00 | 91,00 |
| Port05 | 38 | 59,32 | 23,31 | 39% | 15,00 | 91,00 |
| Port06 | 46 | 43,89 | 25,26 | 58% | 0,0 | 87,00 |
| Port07 | 40 | 25,25 | 23,11 | 92% | 0,0 | 74,00 |
| Port08 | 44 | 50,52 | 21,94 | 43% | 0,0 | 78,00 |

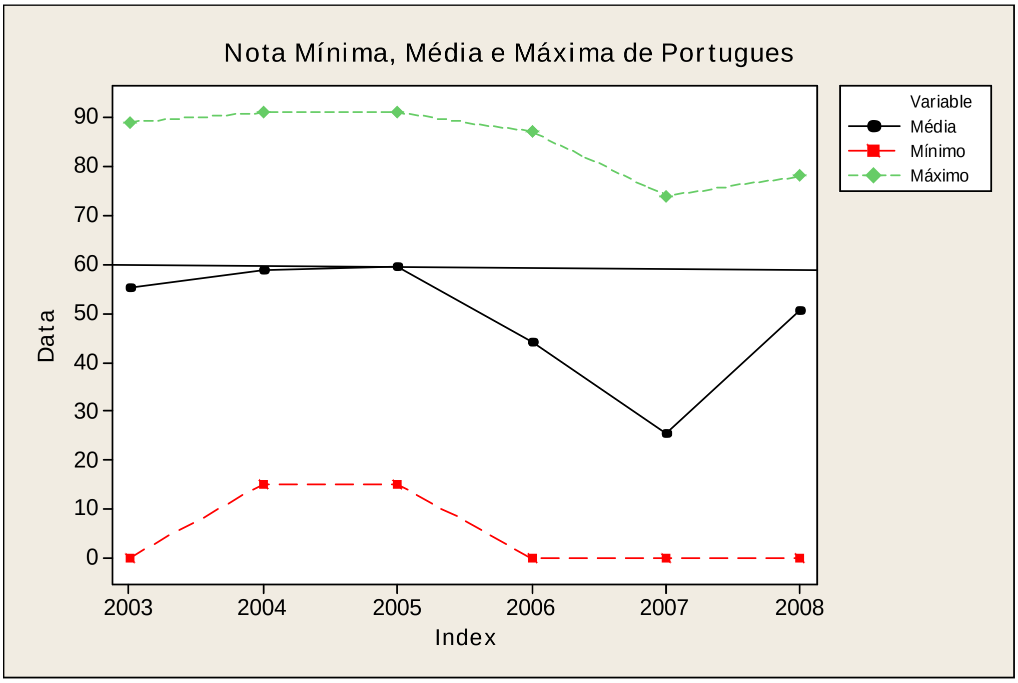



Figura 07: Notas, de português no período de 2003 a 2008

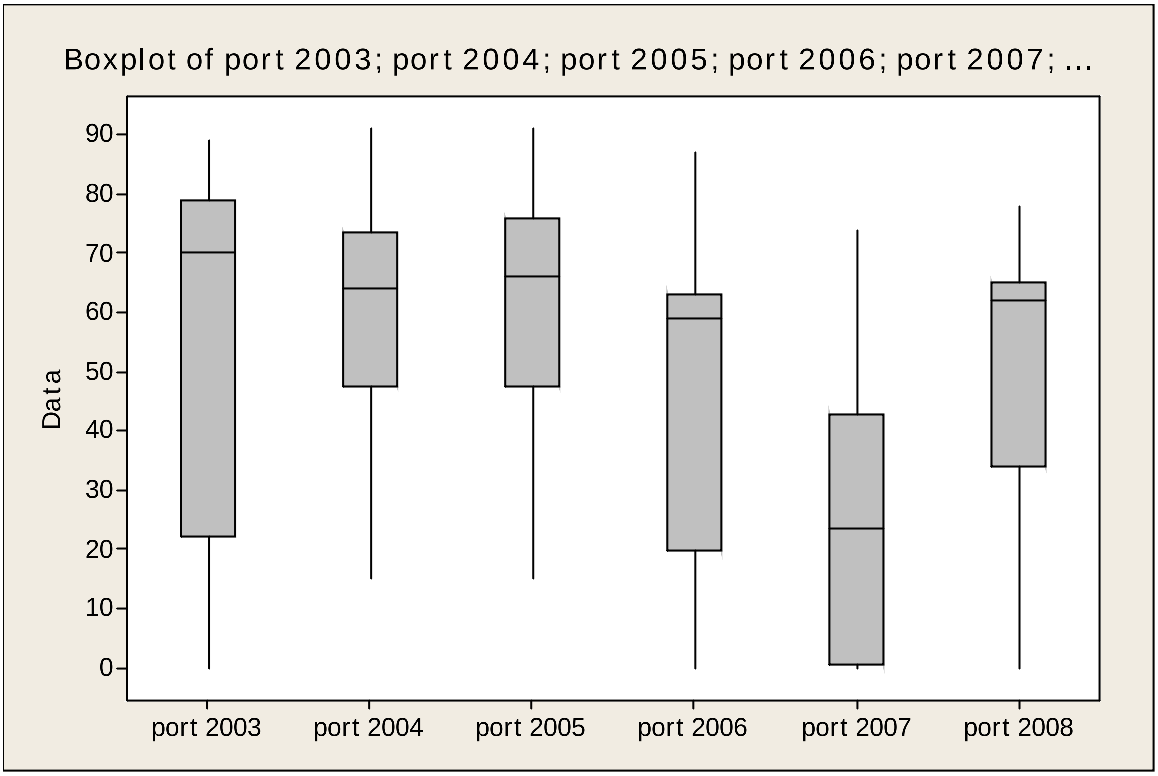



Figura 08: Notas de português no período de 2003 a 2008

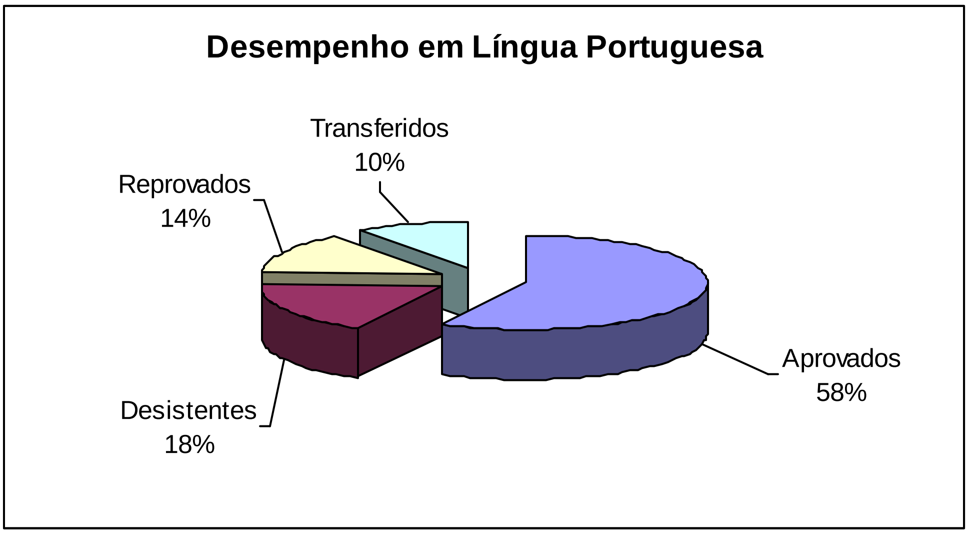



Figura 09: Situação final na disciplina português no período de 2003 a 2008

Agora veremos o aproveitamento, segundo o sexo.

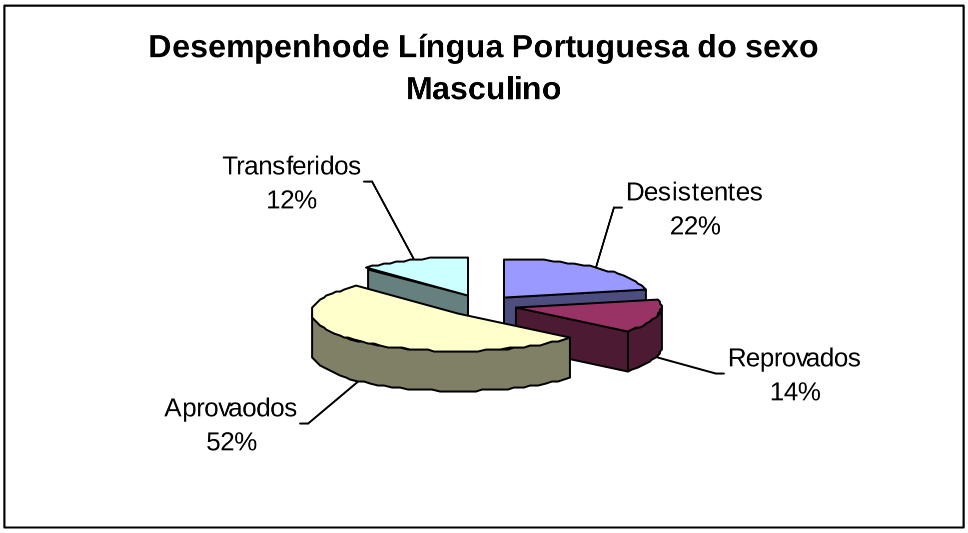



Figura 10: Período de 2003 a 2008

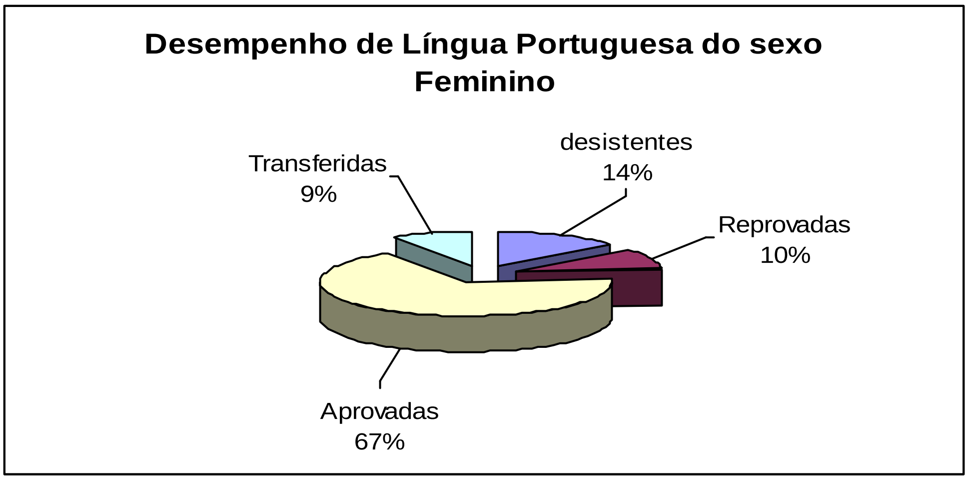



Figura 11: Período de 2003 a 2008

Podemos observar que o rendimento do sexo feminino melhorou consideravelmente em relação ao resultado global de acordo com a figura 10, em relação ao percentual do número de aprovados, considerando também que os valores de reprovação diminuíram seguidos dos desistentes.

Os valores das médias aumentaram em ambos os casos, e os desvios padrões diminuíram de acordo com a figura 11, mostrando assim um melhor desempenho.

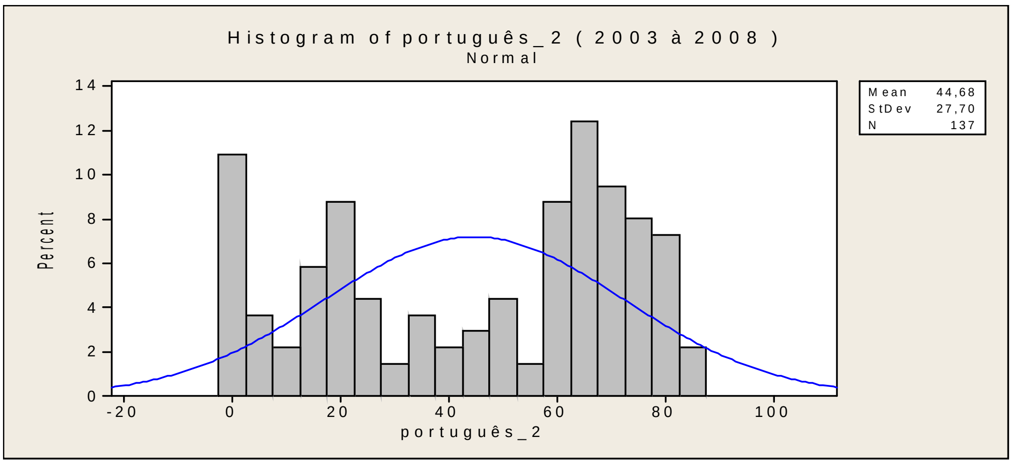



Figura 12: Distribuição do período de 2003 a 2008

.

### 7-Comparação de rendimento em matemática e português

Nesta seção se apresentará uma comparação dos rendimentos em matemática e português

Inicialmente foi avaliado o desempenho em relação ao número de aprovação, desistentes e reprovados entre 2003 a 2008:

Avaliando agora o conteúdo de Língua Portuguesa, podemos observa-se que o resultado em relação à matemática piorou,refletindo o resultado da disciplina de matemática, que já era esperado em virtude da dependência desses dois conteúdos.

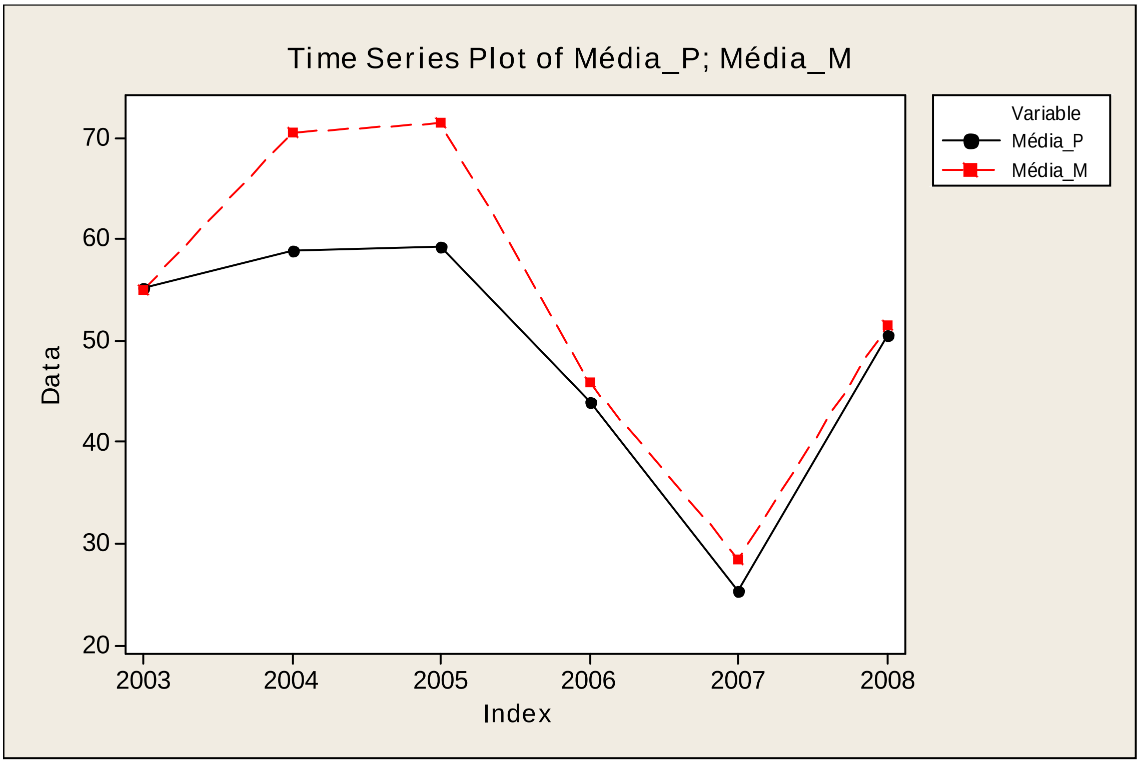



Figura 13: Médias de Matemática e Língua Portuguesa no período de 2003 a 2008.

Verifica-se que os resultados apresentam-se um pouco melhores nos anos 2004 e 2005 no aproveitamento do conteúdo de matemática, pode-se observar também que o ano de 2007 o resultado foi muito ruim nas duas disciplinas.

Iremos testar a hipóteses que a diferença entre as médias é significativamente , para isto testamos se as variâncias são iguais, em anexo os resultados do teste, Há evidencia estatística suficiente para afirmar que os dois grupos provenham de populações com variâncias iguais

### 7.1: Teste da variância

| | Matemática | Português |
|---|---|---|
| Média | 53,6 | 48,9 |
| Desvio Padrão | 29,3 | 27,6 |

**Tabela 4** : Teste da variância entre as médias da disciplina de Língua Portuguesa e Matemática

Ao fazermos o comparativo das duas disciplinas em conjunto podemos reafirmar que uma grande massa dos resultados apresenta valores inferiores à média de 60 pontos como mostra a figura 13.

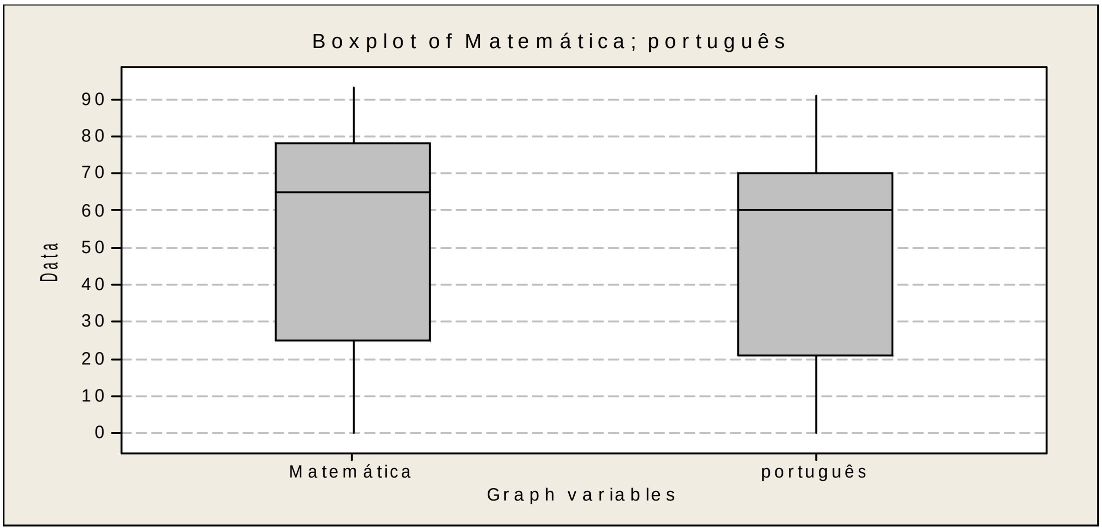



Figura 14: Notas de Matemática e Língua Portuguesa no período de 2003 a 2008.

## 7.2-Relação entre a nota de matemática e português

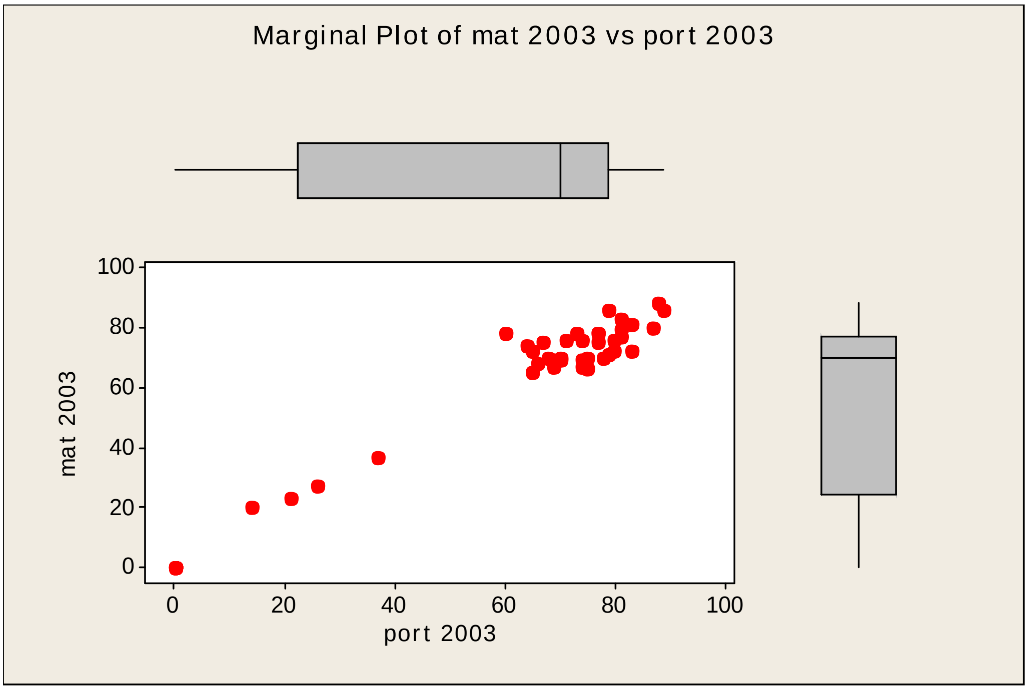



Figura 15:Gráfico de dispersão entre Matemática e Língua Portuguesa (2003)

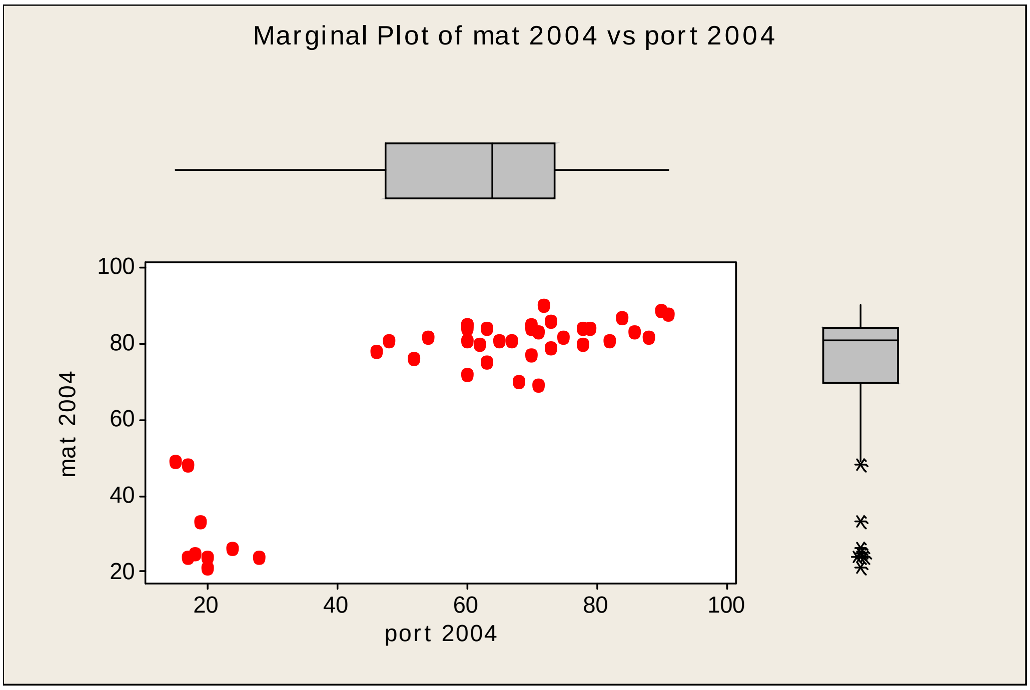



Figura 16:Gráfico de dispersão entre Matemática e Língua Portuguesa (2004)

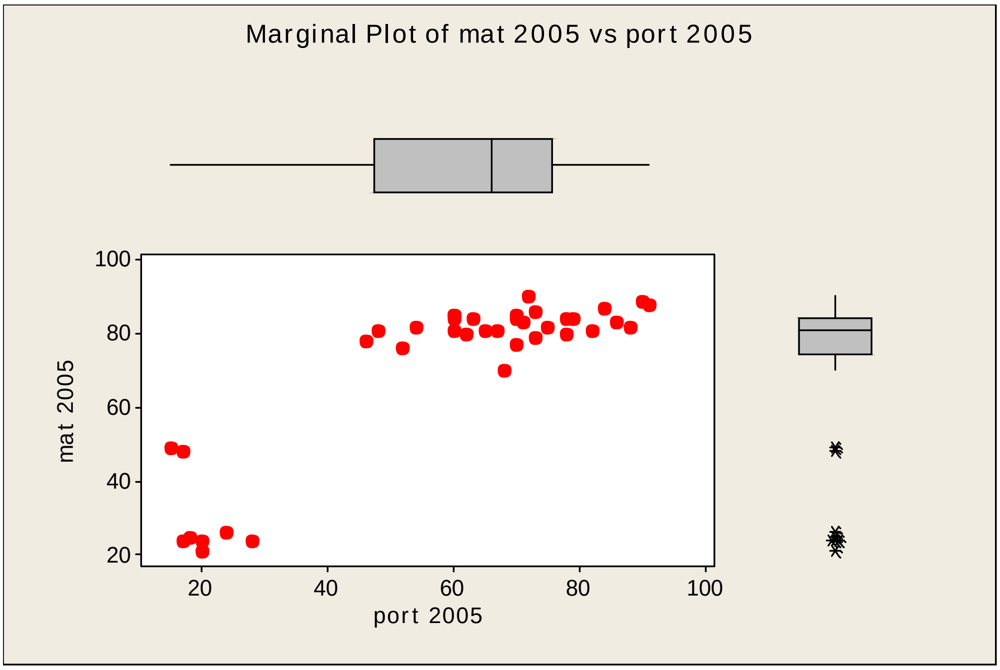



Figura 17:Gráfico de dispersão entre Matemática e Língua Portuguesa (2005)

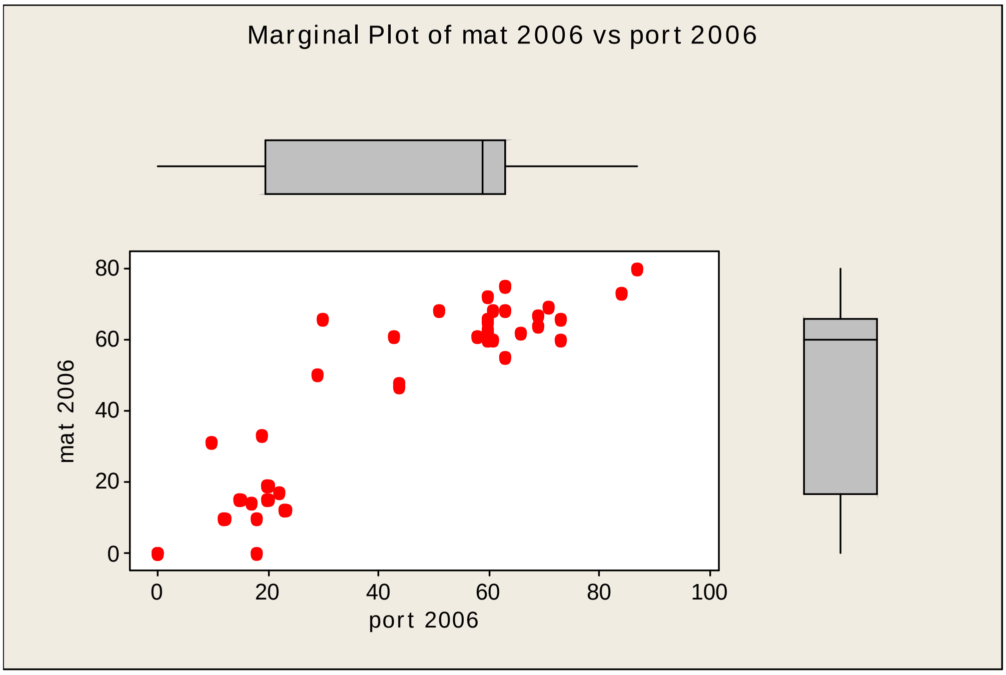



Figura 18:Gráfico de dispersão entre Matemática e Língua Portuguesa (2006)

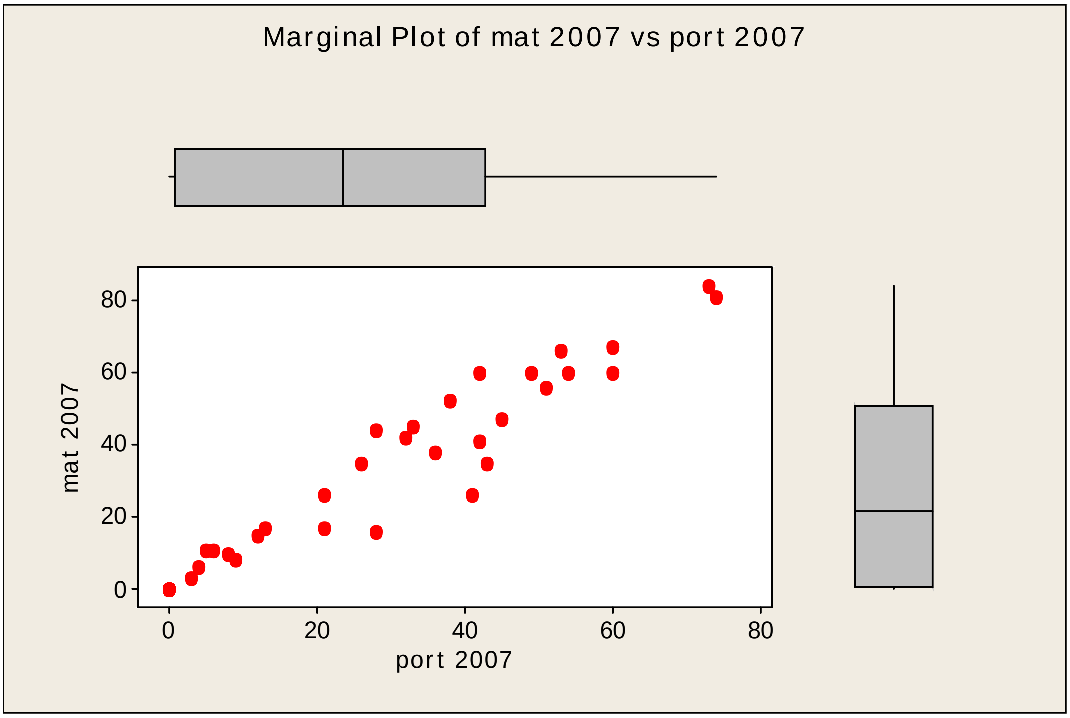



Figura 19:Gráfico de dispersão entre Matemática e Língua Portuguesa (2007)

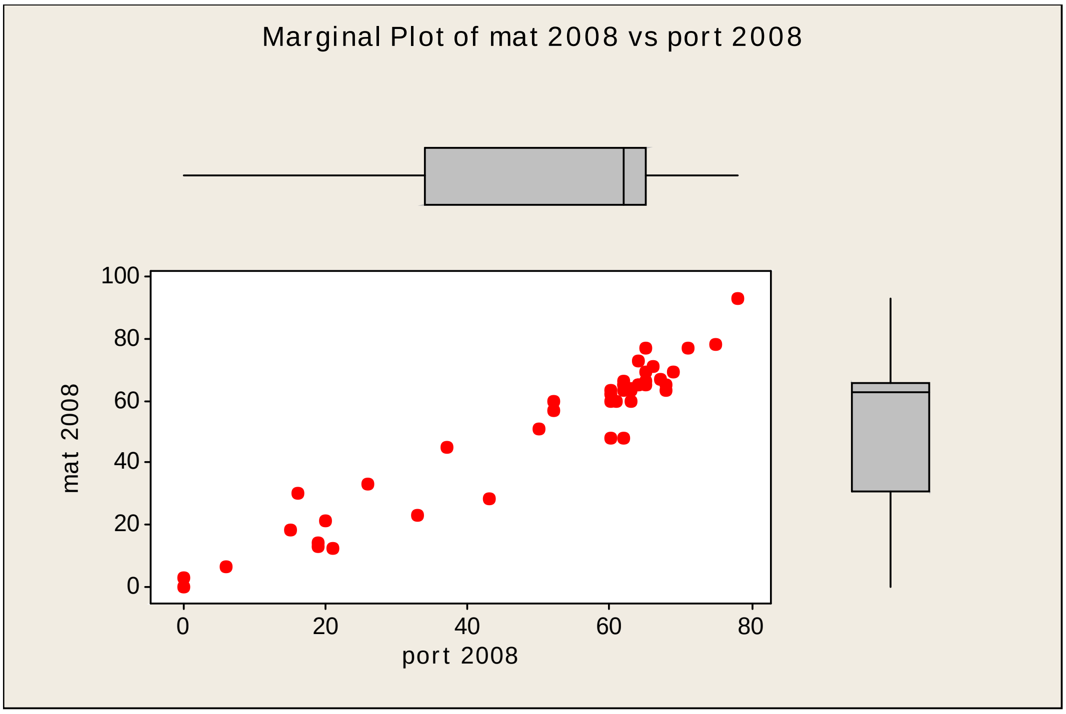



Figura 20:Gráfico de dispersão entre Matemática e Língua Portuguesa (2008)

Podemos observar que os dados apresentam dispersos seguindo um comportamento aproximadamente linear, com correlação positiva.

## 8. Análise de Correlação entre as variáveis

| Ano | Correlação de Pearson Matemática e Português | Valor - P |
|---|---|---|
| 2003 | 0,987 | 0,000 |
| 2004 | 0,877 | 0,000 |
| 2005 | 0,872 | 0,000 |
| 2006 | 0,925 | 0,000 |
| 2007 | 0,969 | 0,000 |
| 2008 | 0,962 | 0,000 |

Tabela 05: Análise de correlação entre as variáveis

Para todos os anos, observa-se uma forte correlação entre as disciplinas. Em todos os casos o valor P da estatística do teste é significativo, ou seja << 5%. Podemos salientar que o rendimento dos conteúdos de matemática e língua portuguesa estão correlacionados, uma vez que há necessidade de leitura e interpretação para o desenvolvimento da matemática.

De acordo com a prática do dia a dia podemos dizer a forte correlação entre os conteúdos de matemática e Língua Portuguesa, tal fato pode ser comprovado a partir das análises dos dados. Esse fato fica claro para os anos de 2004, 2005, porém os demais anos ao testarmos os resíduos verificam-se que os mesmos não são normais, sendo assim vamos propor o teste de Fisher para medirmos a variância entre as duas disciplinas.

## 8.1 Modelo de Regressão

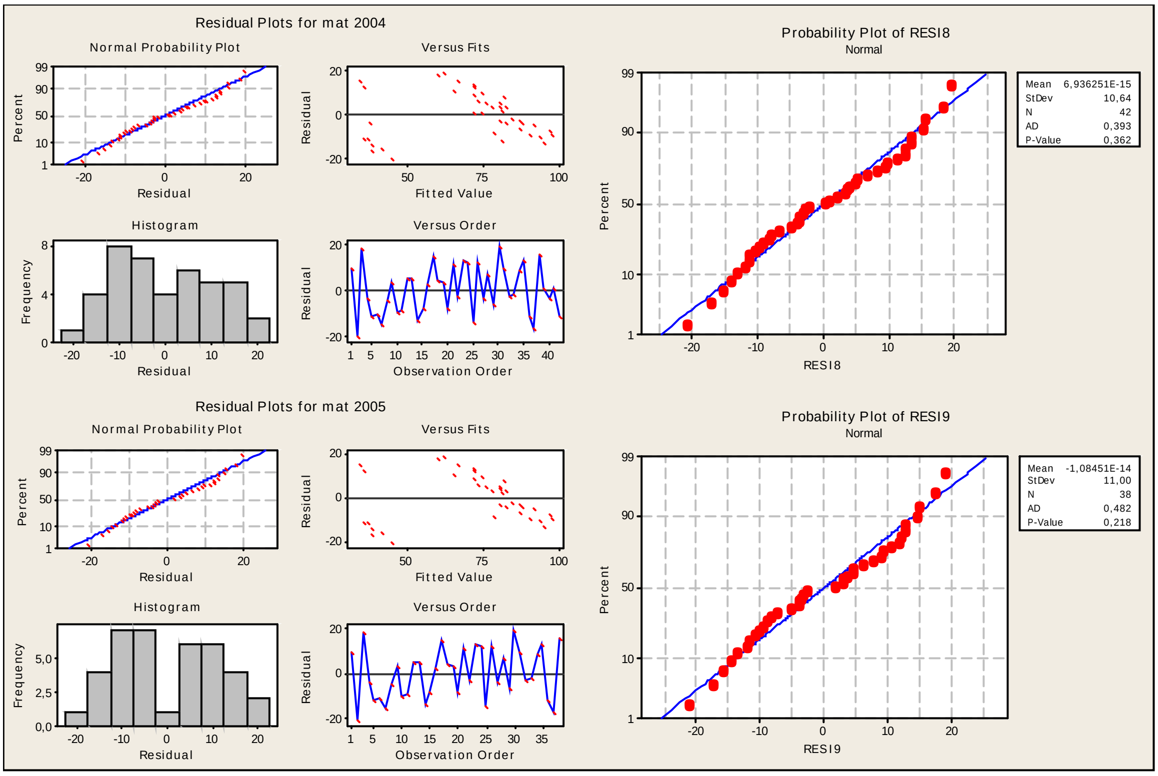



Figura 21:Modelo de regressão para os anos de 2004 e 2005

Os dados apresentam dispersos seguindo um comportamento aproximadamente linear, com correlação positiva.

Pela tabela de análise e variância, conclui que o modelo ajustado é significativo, pois o p-valor >>0,05, ou seja, existe evidência de uma relação linear entre a variável Língua Portuguesa e Matemática.

A análise dos resíduos deletados não detectou nenhum ponto fortemente discrepante, a análise visual dos resíduos, mostra que o comportamento dos mesmos em relação à reta ajustada é visualmente aleatório. O teste de normalidade produziu um p-valor >> 0,05, não rejeitando a hipótese nula e normalidade dos resíduos, portando o modelo e o ajuste satisfaz todas as condições impostas sobre a linearidade e normalidade.

.

Agora fazendo a análise de regressão linear para os anos de 2003, 2006, 2007 e 2008, verificamos:

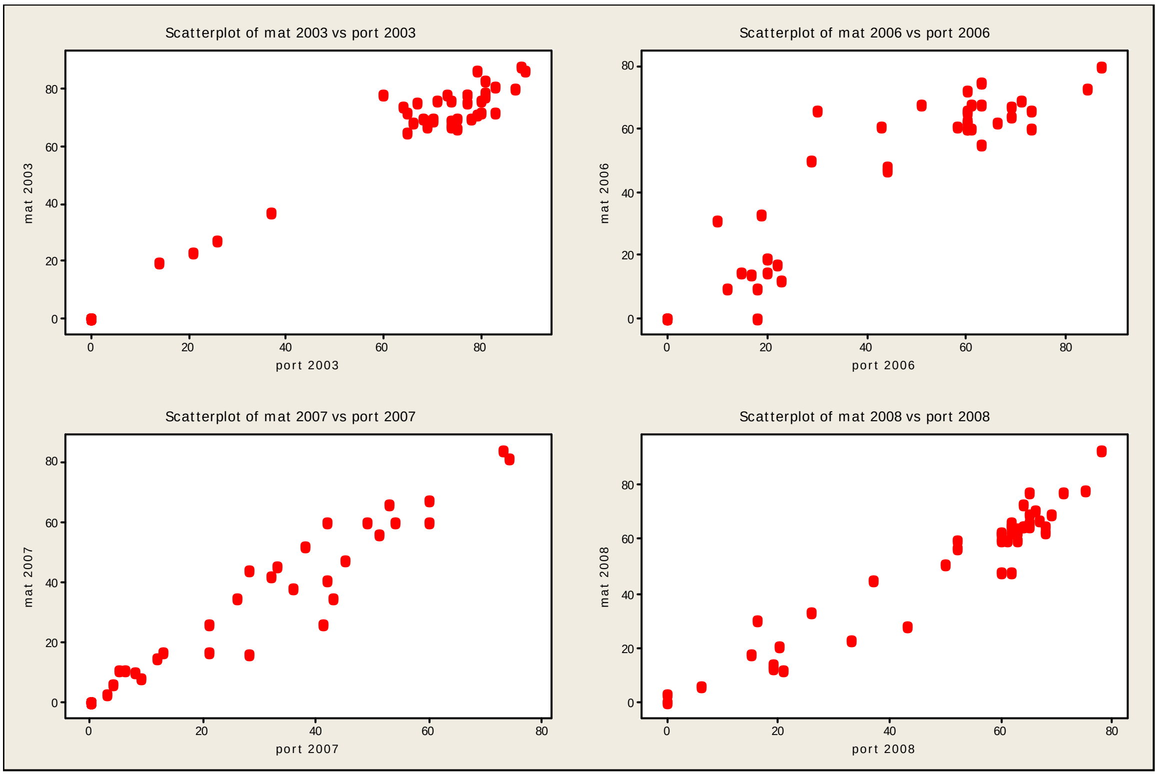



Figura 22: Gráficos de dispersão dos anos de 2003, 2006, 2007 e 2008

Os dados apresentam dispersos seguindo um comportamento aproximadamente linear, com correlação positiva.

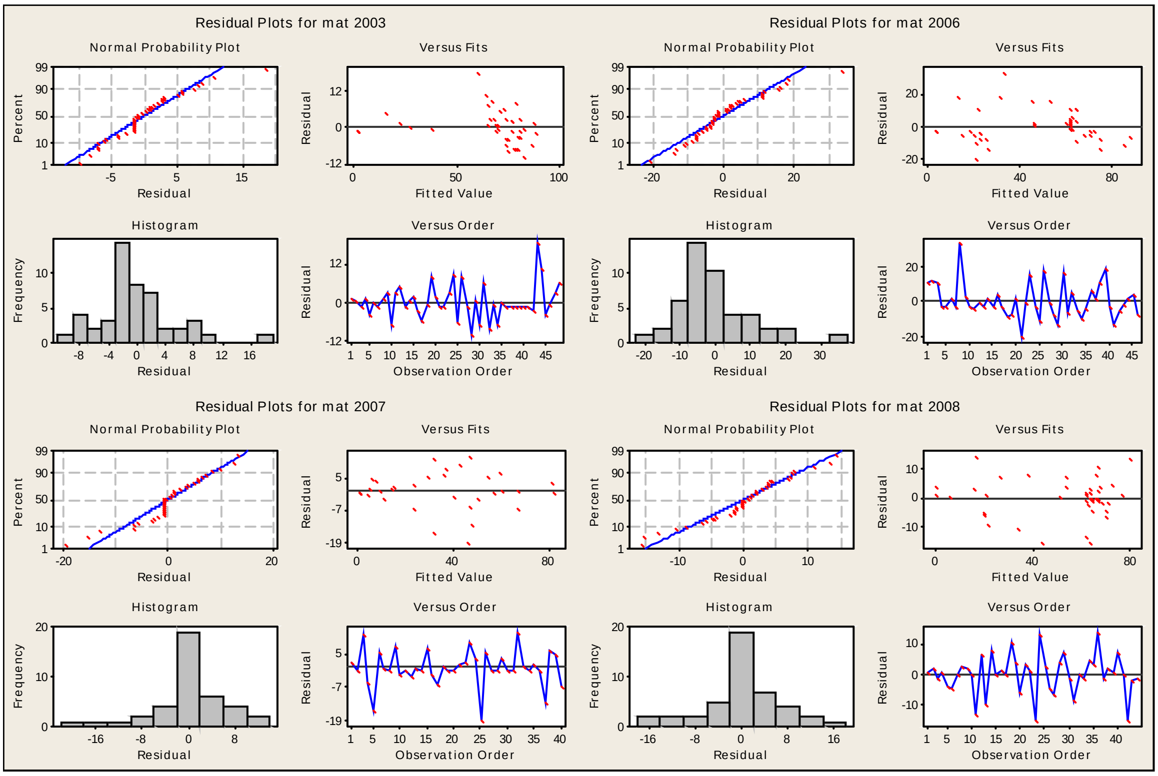



Figura 23: Gráfico contendo o modelo linear para os anos de 2003, 2006, 2007 e 2008.

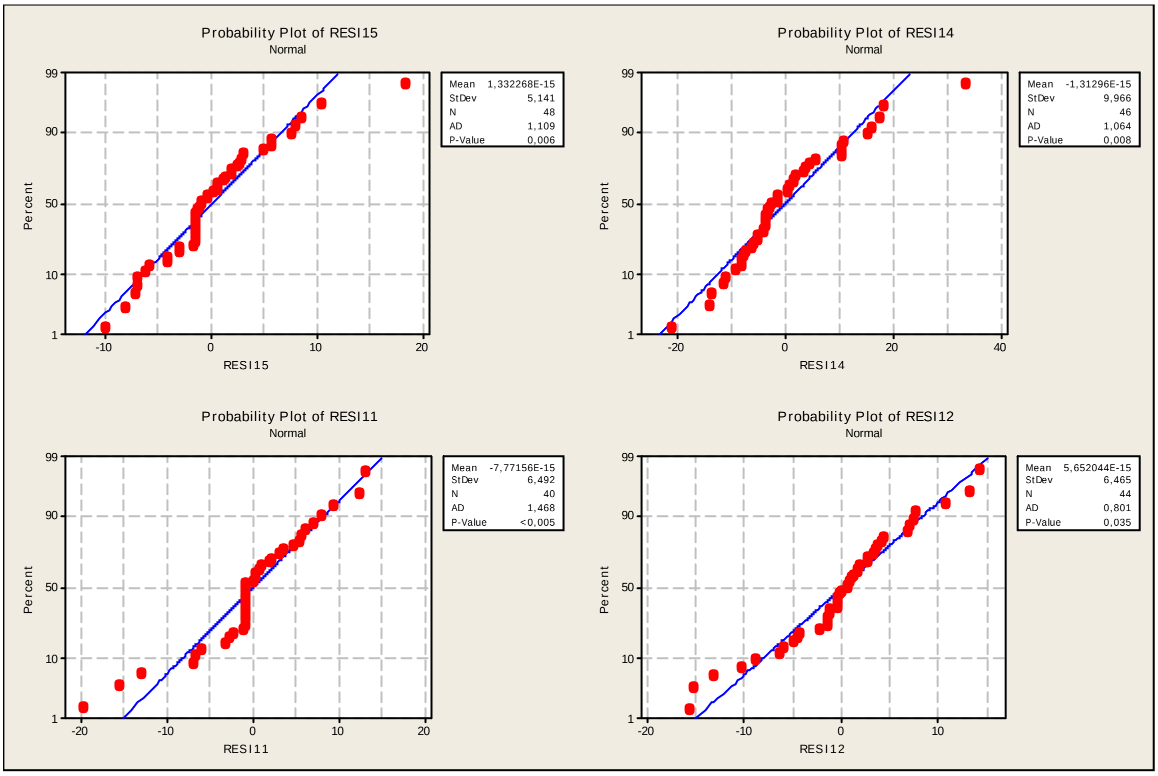



Figura 24: Testes dos resíduos

Pela tabela de análise e variância, conclui que o modelo ajustado é significativo, pois o p-valor >>0,05, ou seja, existe evidência de uma relação linear entre a variável Língua Portuguesa e Matemática.

A análise dos resíduos deletados não detectou nenhum ponto fortemente discrepante, a análise visual dos resíduos, mostra que o comportamento dos mesmos em relação à reta ajustada é visualmente aleatório, porém ao testarmos os resíduos verifica-se que os mesmos não são normais. Sendo assim vamos propor o teste de Fisher para medirmos a variância entre as duas disciplinas.

H0 : Variâncias entre as disciplinas são iguais

H1 : Umas das variáveis é diferente

| **Coef.** | **2003** | **2004** | **2005** | **2006** | **2007** | **2008** |
|---|---|---|---|---|---|---|
| M= | 1,59* | 21,1* | 21,6* | 3,55* | 0,82* | 0,71* |
| N= | 0,968* | 0,841* | 0,84* | 0,963* | 1,1* | 1,03* |
| **R2** | 97,40% | 76,90% | 76,00% | 85,60% | 93,80% | 92,50% |
| **teste norm.** | 0,006 | 0,362 | 0,218 | 0,008 | 0,005 | 0,035 |
| **teste F- p valor** | **0.530** | **0.610** | **0.124** | **0.329** | **0.667** | **0.378** |

**Tabela 06**

*** Os coeficientes significantes diferente de zero.**

De acordo com os resultados encontrados na tabela acima, com os valores de p-valor >>0,05 podemos concluir que as variâncias dos dois grupos são iguais, não rejeitando a hipótese nula.

Agora tendo em vista os resultados encontrados acima, vamos propor uma distribuição quadrática a fim de validarmos os modelos para os anos de 2003, 2006, 2007 e 2008.

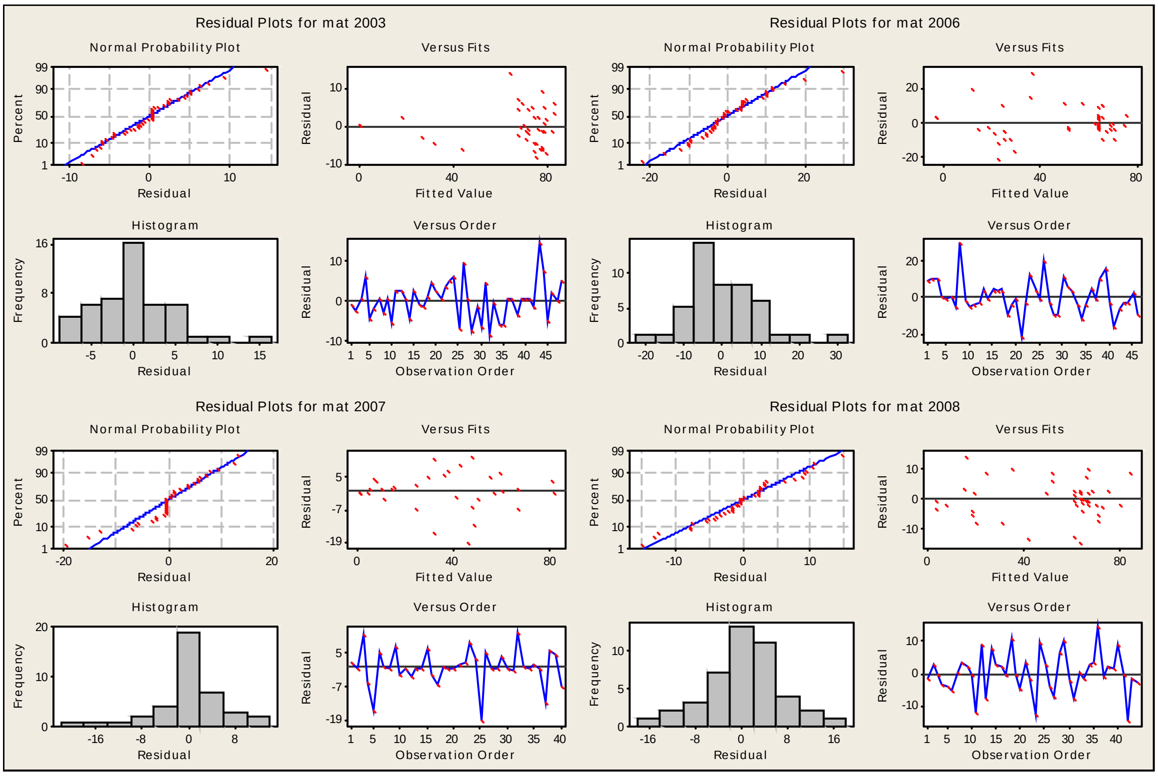



Figura 25: Gráfico contendo a distribuição quadrática para os anos de 2003, 2006, 2007 e 2008.

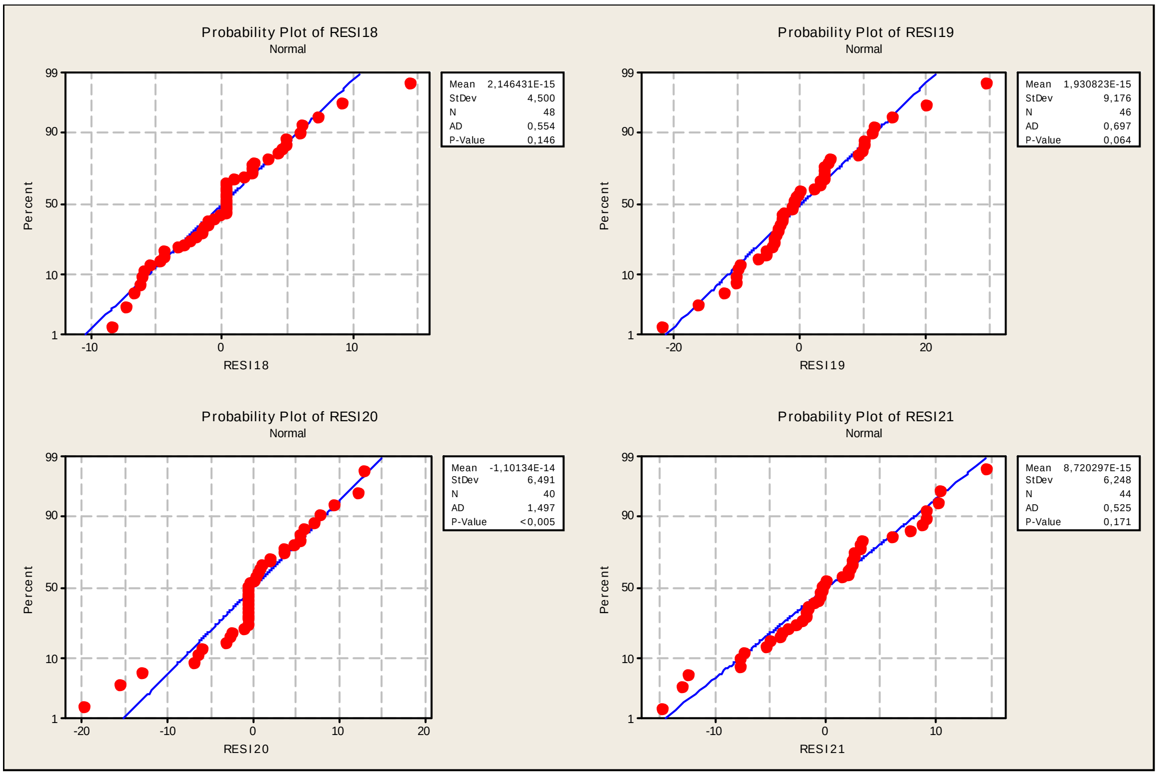



Figura 26: Testes dos resíduos

Pela tabela de análise e variância, conclui que o modelo ajustado é significativo, para os anos de 2003, 2006 e 2008, pois o p-valor >>0,05, ou seja, existe evidência de uma relação entre a variável Língua Portuguesa e Matemática, conseguindo assim validar mais estes três anos seguindo uma distribuição quadrática, ficando apenas o ano de 2007 não validado, visto que foi um ano de altíssima reprovação nos dois conteúdos, assim comprometendo o modelo.

Agora vamos encontrar um modelo seguindo uma distribuição quadrática para o desempenho segundo gênero.

**Modelo ( sexo Masculino):**

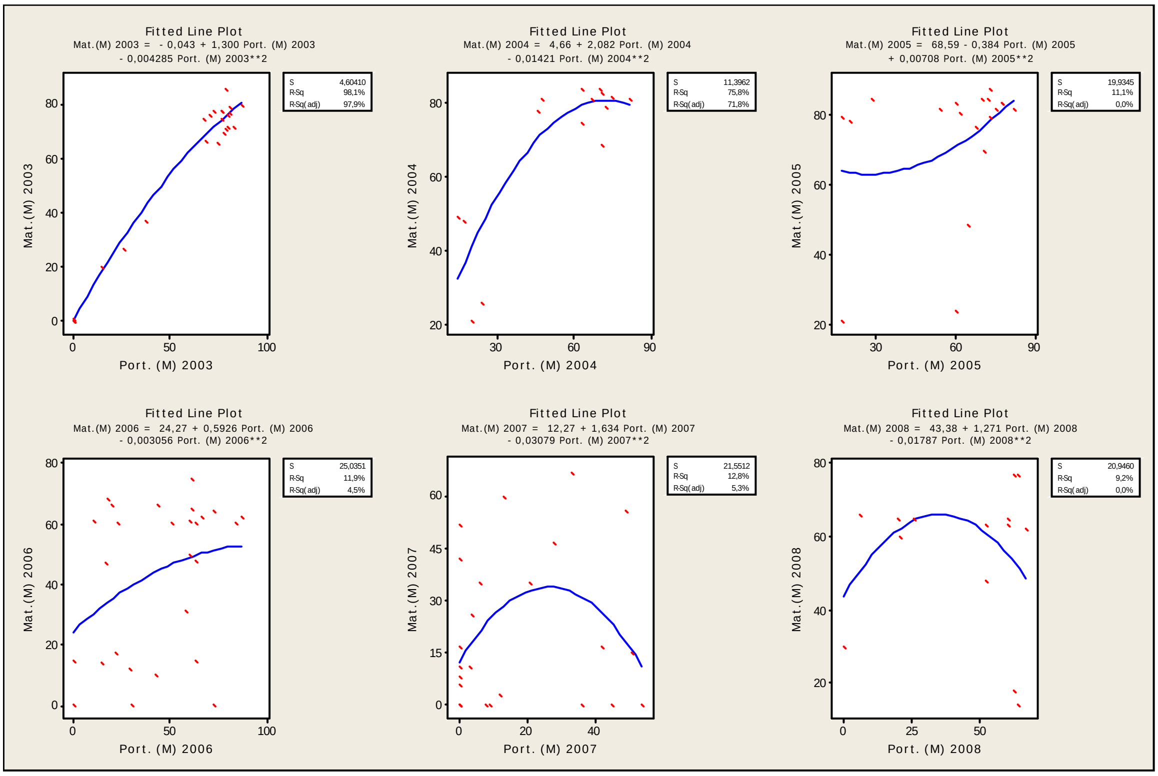



Figura 27: Modelos para os anos de 2003 a 2008 para o sexo masculino

**Sexo Masculino**

| Anos | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 |
|---|---|---|---|---|---|---|
| teste norm. | 0,44 | 0,2 | 0,005 | 0,062 | 0,052 | 0,273 |
| R2 | 98,10% | 75,80% | 11,10% | 11,90% | 12,80% | 9,20% |

Pela tabela de análise e variância, conclui que o modelo ajustado é significativo, para os anos de 2003, 2004, 2006, 2007, e 2008, pois o p-valor >>0,05, ou seja, existe evidência de uma relação entre a variável Língua Portuguesa e Matemática,dentro do grupo separado segundo sexo masculino.

## Modelo ( sexo Feminino):

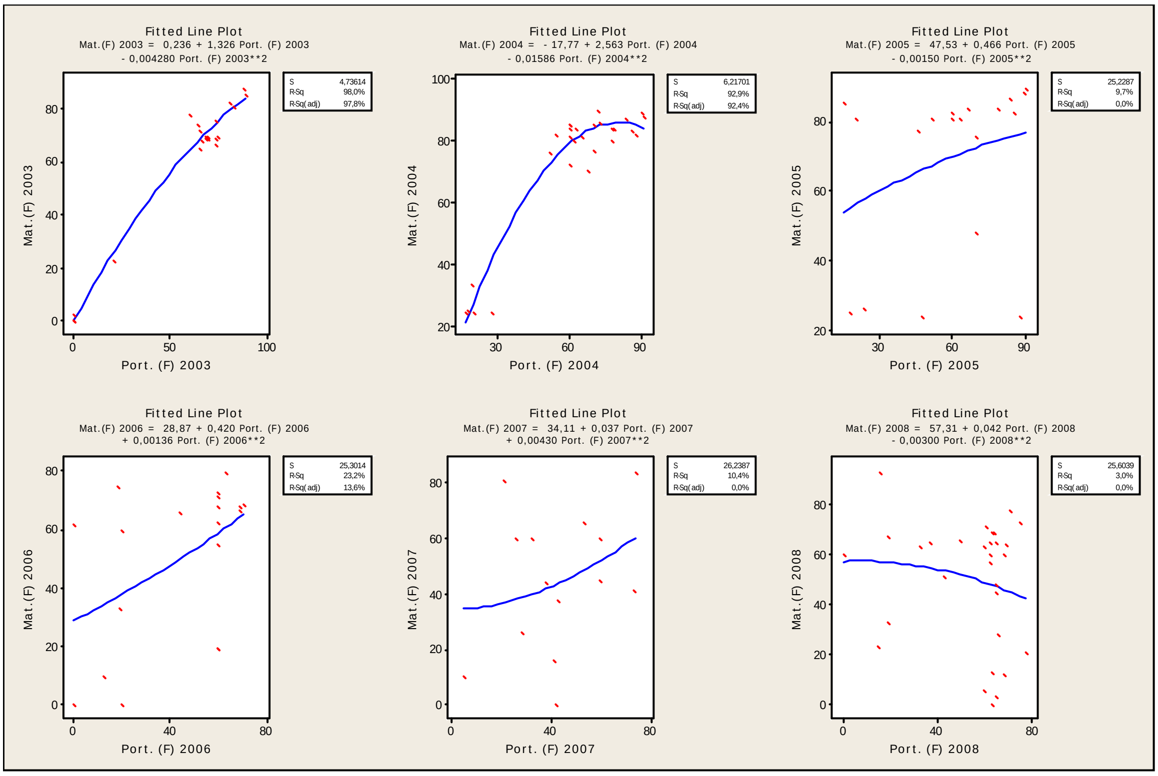



Figura 28: Modelos para os anos de 2003 a 2008 para o sexo feminino

### Sexo Feminino

| **Anos** | **2003** | **2004** | **2005** | **2006** | **2007** | **2008** |
|---|---|---|---|---|---|---|
| **teste norm.** | 0,199 | 0,056 | 0,05 | 0,118 | 0,902 | 0,005 |
| **R2** | 98,00% | 92,90% | 9,70% | 23,20% | 10,40% | 3,00% |

Pela tabela de análise e variância, conclui que o modelo ajustado é significativo, para os anos de 2003, 2004, 2005, 2006, e 2007, pois o p-valor >>0,05, ou seja, existe evidência de uma relação entre a variável Língua Portuguesa e Matemática,dentro do grupo separado segundo sexo masculino.

## 9. CONCLUSÕES.

Ao final dos estudos percebemos a existência da dependência da disciplina de língua Portuguesa influencia diretamente no conteúdo de matemática, onde pode ser comprovado através da regressão linear e do teste de Fisher, visto que as amostras seguem um comportamento independente no que se diz a respeito do indivíduo (aluno). Outro aspecto comprovado é que o desempenho dos alunos segundo o sexo feminino apresenta desempenhos acima do sexo masculino em todos os aspectos estudados de forma positiva. Mas é claro que o desempenho desses alunos depende de uma série de outros fatores que estão direta ou indiretamente ligados, tais como a não capacitação continuada dos professores da rede pública oferecida pelo estado e prefeituras, pouca participação da família no dia a dia escolar, falta de estrutura física e aparelhagem nas escolas, projetos deveriam ser desenvolvidos com a secretaria de educação e professores para amenizar a defasagem de aprendizado de pré requisitos básicos dos alunos do ensino médio para o desenvolvimento da grade curricular CBC (currículo básico comum), esses e outros fatores devem ser abordados em estudos futuros para que se possa garantir um crescimento real da qualidade da educação.

## Bibliografia


GRÁCIO, M. C. C.; GARRUTTI, E. A. A disciplina estatística na área de educação: seleção e organização de conteúdos. Educ. Matem. Ver. - RS, Porto Alegre, v.5, p.12-20,2003a

Montgomery, D. C & Runger, G. C. (2003). Estatística Aplicada e Probabilidade para Engenheiros, 2ª Ed. , LTC Livros Técnicos e Científicos Editora S. A., Rio de Janeiro RJ.

TRIOLA, Mário F. Introdução à Estatística. LTC. 10a edição 2008. 722p

MANN, Prem S. Introdução à Estatística. LTC. 5a edição 2006, 774p

CLARK, Jeffrey & DOWNING, Douglas. Estatística Aplicada. São Paulo, Editora Saraiva, 1998.

COSTA, Sérgio Francisco. Introdução Ilustrada à Estatística. São Paulo, Editora Harbra, 1992.

SILVA, Nelson Pires. Estatística Auto-Explicativa. São Paulo, Editora Érica, 1998

Anexos:

- ## Correlations: port 2003; mat 2003

```
Pearson correlation of port 2003 and mat 2003 = 0,987
P-Value = 0,000

MTB > Correlation  'port 2004' 'mat 2004'.
```

### Correlations: port 2004; mat 2004

```
Pearson correlation of port 2004 and mat 2004 = 0,877
P-Value = 0,000

MTB > Correlation  'port 2005' 'mat 2005'.
```

### Correlations: port 2005; mat 2005

```
Pearson correlation of port 2005 and mat 2005 = 0,872
P-Value = 0,000

MTB > Correlation  'port 2006' 'mat 2006'.
```

### Correlations: port 2006; mat 2006

```
Pearson correlation of port 2006 and mat 2006 = 0,925
P-Value = 0,000

MTB > Correlation  'port 2007' 'mat 2007'.
```

### Correlations: port 2007; mat 2007

```
Pearson correlation of port 2007 and mat 2007 = 0,969
P-Value = 0,000

MTB > Correlation  'port 2008' 'mat 2008'.
```

### Correlations: port 2008; mat 2008

```
Pearson correlation of port 2008 and mat 2008 = 0,962
P-Value = 0,000
```

### Test for Equal Variances: mat 2003 versus port 2003

```
95% Bonferroni confidence intervals for standard deviations

port
2003   N    Lower    StDev    Upper
   0  10        *   0,0000        *
  14   1        *        *        *
  21   1        *        *        *
  26   1        *        *        *
  37   1        *        *        *
  60   1        *        *        *
  64   1        *        *        *
  65   2  1,63718   4,9497  1579,73
  66   1        *        *        *
  67   1        *        *        *
  68   1        *        *        *
  69   2  0,46777   1,4142   451,35
  70   2  0,23388   0,7071   225,68
```

```
  71   1        *        *        *
  73   1        *        *        *
  74   3  1,93068   4,7258    94,46
  75   2  0,93553   2,8284   902,70
  77   3  0,70761   1,7321    34,62
  78   1        *        *        *
  79   2  3,50824  10,6066  3385,13
  80   2  0,93553   2,8284   902,70
  81   3  1,24811   3,0551    61,06
  83   2  2,10494   6,3640  2031,08
  87   1        *        *        *
  88   1        *        *        *
  89   1        *        *        *


Bartlett's Test (Normal Distribution)
Test statistic = 8,04; p-value = 0,530


Levene's Test (Any Continuous Distribution)
Test statistic = 7,09; p-value = 0,000
```

### Test for Equal Variances: mat 2003 versus port 2003

```
MTB > Name c27 "RESI13"
MTB > Regress 'mat 2003' 1  'port 2003';
SUBC>   Residuals 'RESI13';
SUBC>   GFourpack;
SUBC>   RType 1;
SUBC>   Constant;
SUBC>   Brief 2.
```

### Regression Analysis: mat 2003 versus port 2003

```
The regression equation is
mat 2003 = 1,59 + 0,968 port 2003


Predictor     Coef  SE Coef      T      P
Constant     1,585    1,490   1,06  0,293
port 2003  0,96785  0,02335  41,46  0,000


S = 5,19639   R-Sq = 97,4%   R-Sq(adj) = 97,3%


Analysis of Variance

Source          DF     SS     MS        F      P
Regression       1  46410  46410  1718,73  0,000
Residual Error  46   1242     27
Total           47  47652


Unusual Observations

     port
Obs  2003  mat 2003     Fit  SE Fit  Residual  St Resid
 43  60,0    78,000  59,656   0,759    18,344     3,57R
 44  64,0    74,000  63,528   0,778    10,472     2,04R

R denotes an observation with a large standardized residual.
```

- ## Test for Equal Variances: mat 2003 versus port 2003

95% Bonferroni confidence intervals for standard deviations

| port 2003 | N | Lower | StDev | Upper |
|---|---|---|---|---|
| 0 | 10 | * | 0,0000 | * |
| 14 | 1 | * | * | * |
| 21 | 1 | * | * | * |
| 26 | 1 | * | * | * |
| 37 | 1 | * | * | * |
| 60 | 1 | * | * | * |
| 64 | 1 | * | * | * |
| 65 | 2 | 1,63718 | 4,9497 | 1579,73 |
| 66 | 1 | * | * | * |
| 67 | 1 | * | * | * |
| 68 | 1 | * | * | * |
| 69 | 2 | 0,46777 | 1,4142 | 451,35 |
| 70 | 2 | 0,23388 | 0,7071 | 225,68 |
| 71 | 1 | * | * | * |
| 73 | 1 | * | * | * |
| 74 | 3 | 1,93068 | 4,7258 | 94,46 |
| 75 | 2 | 0,93553 | 2,8284 | 902,70 |
| 77 | 3 | 0,70761 | 1,7321 | 34,62 |
| 78 | 1 | * | * | * |
| 79 | 2 | 3,50824 | 10,6066 | 3385,13 |
| 80 | 2 | 0,93553 | 2,8284 | 902,70 |
| 81 | 3 | 1,24811 | 3,0551 | 61,06 |
| 83 | 2 | 2,10494 | 6,3640 | 2031,08 |
| 87 | 1 | * | * | * |
| 88 | 1 | * | * | * |
| 89 | 1 | * | * | * |

Bartlett's Test (Normal Distribution)
Test statistic = 8,04; p-value = 0,530

Levene's Test (Any Continuous Distribution)
Test statistic = 7,09; p-value = 0,000

### Test for Equal Variances: mat 2003 versus port 2003

```
MTB > Vartest 'mat 2004'  'port 2004';
SUBC>   Confidence 95,0.
```

### Test for Equal Variances: mat 2004 versus port 2004

95% Bonferroni confidence intervals for standard deviations

| port 2004 | N | Lower | StDev | Upper |
|---|---|---|---|---|
| 15 | 1 | * | * | * |
| 17 | 2 | 5,74267 | 16,9706 | 4332,96 |
| 18 | 1 | * | * | * |
| 19 | 1 | * | * | * |
| 20 | 2 | 0,71783 | 2,1213 | 541,62 |
| 24 | 1 | * | * | * |
| 28 | 1 | * | * | * |
| 46 | 1 | * | * | * |
| 48 | 1 | * | * | * |
| 52 | 1 | * | * | * |
| 54 | 1 | * | * | * |

| | | | | |
|---|---|---|---|---|
| 60 | 5 | 2,75897 | 5,5045 | 27,32 |
| 62 | 1 | * | * | * |
| 63 | 2 | 2,15350 | 6,3640 | 1624,86 |
| 65 | 1 | * | * | * |
| 67 | 1 | * | * | * |
| 68 | 1 | * | * | * |
| 70 | 3 | 1,81490 | 4,3589 | 77,91 |
| 71 | 2 | 3,34989 | 9,8995 | 2527,56 |
| 72 | 1 | * | * | * |
| 73 | 2 | 1,67495 | 4,9497 | 1263,78 |
| 75 | 1 | * | * | * |
| 78 | 2 | 0,95711 | 2,8284 | 722,16 |
| 79 | 1 | * | * | * |
| 82 | 1 | * | * | * |
| 84 | 1 | * | * | * |
| 86 | 1 | * | * | * |
| 88 | 1 | * | * | * |
| 90 | 1 | * | * | * |
| 91 | 1 | * | * | * |

Bartlett's Test (Normal Distribution)
Test statistic = 5,41; p-value = 0,610

Levene's Test (Any Continuous Distribution)
Test statistic = 2,35; p-value = 0,092

## Test for Equal Variances: mat 2004 versus port 2004

```
MTB > Vartest 'mat 2005'  'port 2005';
SUBC>   Confidence 95,0.
```

## Test for Equal Variances: mat 2005 versus port 2005

95% Bonferroni confidence intervals for standard deviations

| port 2005 | N | Lower | StDev | Upper |
|---|---|---|---|---|
| 15 | 1 | * | * | * |
| 17 | 2 | 5,92287 | 16,9706 | 3249,72 |
| 18 | 1 | * | * | * |
| 20 | 2 | 0,74036 | 2,1213 | 406,21 |
| 24 | 1 | * | * | * |
| 28 | 1 | * | * | * |
| 46 | 1 | * | * | * |
| 48 | 1 | * | * | * |
| 52 | 1 | * | * | * |
| 54 | 1 | * | * | * |
| 60 | 4 | 0,90145 | 1,8930 | 13,02 |
| 62 | 1 | * | * | * |
| 63 | 1 | * | * | * |
| 65 | 1 | * | * | * |
| 67 | 1 | * | * | * |
| 68 | 1 | * | * | * |
| 70 | 3 | 1,86192 | 4,3589 | 67,46 |
| 71 | 1 | * | * | * |
| 72 | 1 | * | * | * |
| 73 | 2 | 1,72750 | 4,9497 | 947,83 |
| 75 | 1 | * | * | * |
| 78 | 2 | 0,98714 | 2,8284 | 541,62 |
| 79 | 1 | * | * | * |
| 82 | 1 | * | * | * |
| 84 | 1 | * | * | * |
| 86 | 1 | * | * | * |
| 88 | 1 | * | * | * |

```
  90  1        *        *        *
  91  1        *        *        *


Bartlett's Test (Normal Distribution)
Test statistic = 8,65; p-value = 0,124


Levene's Test (Any Continuous Distribution)
Test statistic = 9,11; p-value = 0,003
```

### Test for Equal Variances: mat 2005 versus port 2005

```
MTB > Vartest 'mat 2006'  'port 2006';
SUBC>   Confidence 95,0.
```

### Test for Equal Variances: mat 2006 versus port 2006

```
95% Bonferroni confidence intervals for standard deviations

port
2006  N    Lower    StDev    Upper
   0  4        *  0,00000        *
  10  1        *        *        *
  12  1        *        *        *
  15  1        *        *        *
  17  1        *        *        *
  18  2  2,39278  7,07107  1805,40
  19  1        *        *        *
  20  3  0,96156  2,30940    41,28
  22  1        *        *        *
  23  1        *        *        *
  29  1        *        *        *
  30  1        *        *        *
  43  1        *        *        *
  44  2  0,23928  0,70711   180,54
  51  1        *        *        *
  58  1        *        *        *
  60  9  2,29351  3,90512    10,19
  61  2  1,91422  5,65685  1444,32
  63  4  4,38952  9,42956    71,47
  66  1        *        *        *
  69  2  0,71783  2,12132   541,62
  71  1        *        *        *
  73  2  1,43567  4,24264  1083,24
  84  1        *        *        *
  87  1        *        *        *


Bartlett's Test (Normal Distribution)
Test statistic = 8,05; p-value = 0,329


Levene's Test (Any Continuous Distribution)
Test statistic = 1,58; p-value = 0,192
```

### Test for Equal Variances: mat 2006 versus port 2006

```
MTB > Vartest 'mat 2007'  'port 2007';
SUBC>   Confidence 95,0.
```

### Test for Equal Variances: mat 2007 versus port 2007

```
95% Bonferroni confidence intervals for standard deviations
```

```
port
2007   N     Lower    StDev    Upper
   0  10         *   0,0000        *
   3   1         *        *        *
   4   1         *        *        *
   5   1         *        *        *
   6   1         *        *        *
   8   1         *        *        *
   9   1         *        *        *
  12   1         *        *        *
  13   1         *        *        *
  21   2   2,32740   6,3640   812,42
  26   1         *        *        *
  28   2   7,24079  19,7990  2527,54
  32   1         *        *        *
  33   1         *        *        *
  36   1         *        *        *
  38   1         *        *        *
  41   1         *        *        *
  42   2   4,91339  13,4350  1715,12
  43   1         *        *        *
  45   1         *        *        *
  49   1         *        *        *
  51   1         *        *        *
  53   1         *        *        *
  54   1         *        *        *
  60   2   1,81020   4,9497   631,89
  73   1         *        *        *
  74   1         *        *        *


Bartlett's Test (Normal Distribution)
Test statistic = 1,57; p-value = 0,667

* NOTE * Levene's test cannot be computed for these data.
```

### Test for Equal Variances: mat 2007 versus port 2007

```
MTB > Vartest 'mat 2008'  'port 2008';
SUBC>   Confidence 95,0.
```

### Test for Equal Variances: mat 2008 versus port 2008

```
95% Bonferroni confidence intervals for standard deviations

port
2008  N    Lower    StDev    Upper
   0  2  0,70916  2,12132  609,324
   6  1        *        *        *
  15  1        *        *        *
  16  1        *        *        *
  19  2  0,23639  0,70711  203,108
  20  1        *        *        *
  21  1        *        *        *
  26  1        *        *        *
  33  1        *        *        *
  37  1        *        *        *
  43  1        *        *        *
  50  1        *        *        *
  52  2  0,70916  2,12132  609,324
  60  4  3,20455  6,94622   54,776
  61  1        *        *        *
  62  5  3,76287  7,56968   38,716
  63  3  0,85802  2,08167   39,469
  64  3  1,90377  4,61880   87,575
  65  4  2,50923  5,43906   42,891
```

```
  66  1        *        *        *
  67  1        *        *        *
  68  2  0,47277  1,41421  406,216
  69  1        *        *        *
  71  1        *        *        *
  75  1        *        *        *
  78  1        *        *        *
```

```
Bartlett's Test (Normal Distribution)
Test statistic = 8,59; p-value = 0,378
```

```
Levene's Test (Any Continuous Distribution)
Test statistic = 0,27; p-value = 0,967
```

- ## Test and CI for Two Proportions

```
Sample   X    N  Sample p
1       53  100  0,530000
2       40  120  0,333333
```

```
Difference = p (1) - p (2)
Estimate for difference:  0,196667
95% CI for difference:  (0,0675044; 0,325829)
Test for difference = 0 (vs not = 0):  Z = 2,98  P-Value = 0,003
```

### Test for Equal Variances

```
95% Bonferroni confidence intervals for standard deviations

Sample    N    Lower    StDev    Upper
     1  258  26,6394  29,2800  32,4775
     2  258  25,0928  27,5801  30,5919
```

```
F-Test (Normal Distribution)
Test statistic = 1,13; p-value = 0,338
```

### Two-Sample T-Test and CI

```
Sample    N  Mean  StDev  SE Mean
1       258  53,6   29,3      1,8
2       258  48,9   27,6      1,7
```

```
Difference = mu (1) - mu (2)
Estimate for difference:  4,67
95% CI for difference:  (-0,25; 9,59)
T-Test of difference = 0 (vs not =): T-Value = 1,86  P-Value = 0,063  DF = 514
Both use Pooled StDev = 28,4427
```

- ## Polynomial Regression Analysis: mat 2003 versus port 2003

```
The regression equation is
mat 2003 = - 0,247 + 1,343 port 2003 - 0,004632 port 2003**2


S = 4,59906   R-Sq = 98,0%   R-Sq(adj) = 97,9%


Analysis of Variance

Source      DF       SS       MS        F      P
Regression   2  46700,1  23350,1  1103,95  0,000
Error       45    951,8     21,2
Total       47  47651,9


Sequential Analysis of Variance

Source     DF       SS        F      P
Linear      1  46409,8  1718,73  0,000
Quadratic   1    290,3    13,72  0,001
```

### Fitted Line: mat 2003 versus port 2003

### Residual Plots for mat 2003

```
MTB > NormTest 'RESI18'.
```

### Probability Plot of RESI18

```
MTB > Name c33 "RESI19"
MTB > Fitline 'mat 2006' 'port 2006';
SUBC>   Poly 2;
SUBC>   GFourpack;
SUBC>   RType 1;
SUBC>   Confidence 95,0;
SUBC>   Resid  'RESI19'.
```

### Polynomial Regression Analysis: mat 2006 versus port 2006

```
The regression equation is
mat 2006 = - 3,629 + 1,569 port 2006 - 0,007610 port 2006**2


S = 9,38736   R-Sq = 87,8%   R-Sq(adj) = 87,3%


Analysis of Variance

Source      DF       SS       MS       F      P
Regression   2  27315,3  13657,7  154,98  0,000
Error       43   3789,3     88,1
Total       45  31104,6


Sequential Analysis of Variance

Source     DF       SS       F      P
Linear      1  26634,7  262,18  0,000
Quadratic   1    680,6    7,72  0,008
```

### Fitted Line: mat 2006 versus port 2006

## Residual Plots for mat 2006

```
MTB > NormTest 'RESI19'.
```

## Probability Plot of RESI19

```
MTB > Name c34 "RESI20"
MTB > Fitline 'mat 2007' 'port 2007';
SUBC>   Poly 2;
SUBC>   GFourpack;
SUBC>   RType 1;
SUBC>   Confidence 95,0;
SUBC>   Resid  'RESI20'.
```

## Polynomial Regression Analysis: mat 2007 versus port 2007

```
The regression equation is
mat 2007 = 0,721 + 1,111 port 2007 - 0,000268 port 2007**2


S = 6,66396   R-Sq = 93,8%   R-Sq(adj) = 93,5%


Analysis of Variance

Source      DF       SS       MS       F      P
Regression   2  24996,9  12498,4  281,44  0,000
Error       37   1643,1     44,4
Total       39  26640,0


Sequential Analysis of Variance

Source     DF       SS       F      P
Linear      1  24996,3  577,87  0,000
Quadratic   1      0,6    0,01  0,907
```

## Fitted Line: mat 2007 versus port 2007

## Residual Plots for mat 2007

```
MTB > NormTest 'RESI20'.
```

## Probability Plot of RESI20

```
MTB > Name c35 "RESI21"
MTB > Fitline 'mat 2008' 'port 2008';
SUBC>   Poly 2;
SUBC>   GFourpack;
SUBC>   RType 1;
SUBC>   Confidence 95,0;
SUBC>   Resid  'RESI21'.
```

## Polynomial Regression Analysis: mat 2008 versus port 2008

```
The regression equation is
mat 2008 = 3,536 + 0,6801 port 2008 + 0,004484 port 2008**2


S = 6,39869   R-Sq = 93,0%   R-Sq(adj) = 92,6%
```

```
Analysis of Variance

Source      DF       SS       MS       F      P
Regression   2  22182,2  11091,1  270,89  0,000
Error       41   1678,7     40,9
Total       43  23860,9


Sequential Analysis of Variance

Source     DF       SS       F      P
Linear      1  22063,9  515,68  0,000
Quadratic   1    118,3    2,89  0,097
```

## Fitted Line: mat 2008 versus port 2008

## Residual Plots for mat 2008

```
MTB > NormTest 'RESI21'.
```

## Probability Plot of RESI21