

# Relatório 02

## ANÁLISE DE COMPONENTES PRINCIPAIS - ACP

### “Principal Components Analysis - PCA”

Cristiano Lopes Moreira

Matrícula: 119103-0

Aluno		RA/Matrícula		Professor		Tipo	
Cristiano Lopes Moreira		119103-0		Dr Reinaldo Bianchi		Relatório de implementação	
Data	Versão	Turma		Nome do arquivo			Página
10/11/2019	1	2º. Semestre de 2019		PEL_208_Relatório_02_Cristiano_Moreira.doc			1 (21)

## Sumário

1.	Introdução . . . . .	3
2.	Desenvolvimento teórico . . . . .	3
2.1.	Método para obtenção do PCA . . . . .	4
2.1.1.	Etapa 2 . . . . .	4
2.1.2.	Etapa 3 . . . . .	5
2.1.3.	Etapa 4 . . . . .	5
2.1.4.	Etapa 5 . . . . .	6
2.1.5.	Novo conjunto de dados . . . . .	7
3.	Proposta de implementação . . . . .	7
3.1.	Algoritmo de Análise de Componentes Principais PCA: . . . . .	7
3.2.	Pseudocódigos . . . . .	7
4.	Experimentação e Resultados . . . . .	8
4.1.	Censo Americano . . . . .	10
4.2.	Pressão x Temperatura de Ebulição da água nos Alpes . . . . .	14
4.3.	Livros, palestras e nota . . . . .	18
5.	Conclusão . . . . .	21
6.	Referências . . . . .	21

Aluno		RA/Matrícula		Professor		Tipo	
Cristiano Lopes Moreira		119103-0		Dr Reinaldo Bianchi		Relatório de implementação	
Data	Versão	Turma		Nome do arquivo			Página
10/11/2019	1	2º. Semestre de 2019		PEL_208_Relatório_02_Cristiano_Moreira.doc			2 (21)

## 1. Introdução

O PCA é uma técnica utilizada para de identificar padrões nos dados e expressá-los de maneira a destacar suas semelhanças e diferenças. Como os padrões de dados podem ser difíceis de encontrar em dados de alta dimensão, principalmente naqueles em que não é possível uma representação gráfica, o PCA é uma ferramenta poderosa para analisar dados sendo utilizado para reduzir o número de dimensões em componentes principais (de maior variação dos dados), sem muita perda de informações, e algumas vezes utilizada na compactação de imagens.

## 2. Desenvolvimento teórico

O PCA é um método matemático, que utiliza de conceitos de estatística: variância, desvio padrão, covariância, autovetores e autovalores; tem por finalidade básica, a análise dos dados usados visando sua redução, eliminação de sobreposições e a escolha das formas mais representativas de dados a partir de combinações lineares das variáveis originais.

Por este método é possível identificar uma componente principal, eixo ou agrupamento que melhor representa a distribuição dos dados, (Linha vermelha na Figura 1) e a componente secundária, ortogonal à componente principal (linha verde).

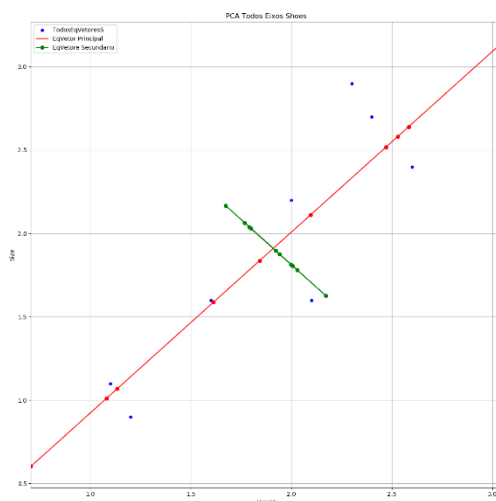


Fig1: Análise de comp. Principal Sapatos [tamanho x altura]

Aluno		RA/Matrícula		Professor	Tipo
Cristiano Lopes Moreira		119103-0		Dr Reinaldo Bianchi	Relatório de implementação
Data	Versão	Turma		Nome do arquivo	Página
10/11/2019	1	2º. Semestre de 2019		PEL_208_Relatório_02_Cristiano_Moreira.doc	3 (21)

## 2.1. Método para obtenção do PCA

6 etapas são necessárias para realizar uma Análise dos Componentes Principais – PCA.

1 Etapa: Obter conjunto de dados.

2 Etapa: Subtrair a média.

3 Etapa: Calcular a matriz de covariância.

4 Etapa: Calcular os autovetores e autovalores da matriz de covariância.

5 Etapa: Escolhendo componentes e formando um vetor de recursos.

6 Etapa: Definir o novo conjunto de dados.

O autovetor com o maior autovalor associado, corresponde à componente principal do conjunto de dados usados, essa componente é a mais significativa na dimensão dos dados.

### 2.1.1. Etapa 2

**Média aritmética**, ou simplesmente média, é a soma do total de valores de uma variável dividida pelo número total de observações

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (1)$$

em que  $n$  é o número total de observações no conjunto de dados  $X_i$ , para  $i=1, \dots, n$ , representando cada um dos valores de  $x$  (FÁVERO et al., 2009).

**Variância** é a medida de dispersão dos dados em torno da média (FÁVERO et al., 2009).

Aluno		RA/Matrícula		Professor		Tipo	
Cristiano Lopes Moreira		119103-0		Dr Reinaldo Bianchi		Relatório de implementação	
Data	Versão	Turma		Nome do arquivo			Página
10/11/2019	1	2º. Semestre de 2019		PEL_208_Relatório_02_Cristiano_Moreira.doc			4 (21)

Para população 
$$\sigma^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n} \quad (2)$$

Para amostras 
$$s^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n - 1} \quad (3)$$

Na segunda etapa ser calculado a média de cada variável do conjunto de dados e realizado um novo conjunto de dados normalizado pela média:

$$(X_i - \bar{X}) \quad (4)$$

### 2.1.2. Etapa 3

**Covariância** é a média do grau de interdependência ou inter-relação numérica linear entre duas variáveis, é o elemento fora da diagonal,  $\varepsilon(X_i - \mu_i)(X_j - \mu_j)$ , a covariância de  $X_i$  e  $X_j, i \neq j$  denota matriz de covariância por  $\Sigma$ . (WILBUR, 1971)

$$\Sigma(x, y) = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{N - 1} \quad (7)$$

Na etapa 3 deve ser calculada a covariância entre todas as variáveis do conjunto de dados para a criação da matriz de covariância  $\Sigma$ .

### 2.1.3. Etapa 4

Autovalores, eigenvalues  $\Lambda$ , mostra a variância total explicada por cada dimensão, quanto maior o valor do eigenvalue maior é a variância explicada por sua dimensão (FÁVERO et al., 2009).

$$\text{Det} |(\Sigma - \Lambda I)| = 0 \quad (8)$$

Aluno		RA/Matrícula		Professor		Tipo	
Cristiano Lopes Moreira		119103-0		Dr Reinaldo Bianchi		Relatório de implementação	
Data	Versão	Turma		Nome do arquivo			Página
10/11/2019	1	2º. Semestre de 2019		PEL_208_Relatório_02_Cristiano_Moreira.doc			5 (21)

Na etapa 4 deve ser calculado os autovalores, que são as raízes do determinante da diferença da matriz de covariância  $\Sigma$  e a matriz identidade multiplicada pelo eigenvalues.

Para calcular o determinante, um conjunto de equações lineares deve ser desenvolvido. Os casos de equações de segunda ordem podem ser calculados pela fórmula de Bhaskara

$$x = \frac{-b \pm \sqrt{\Delta}}{2a} \quad (8)$$

$$\Delta = b^2 - 4ac \quad (9)$$

Para equações polinomiais de ordens superiores deve ser utilizado a transformação de Jacobi, o método consiste em uma sequência de transformação de similaridade ortogonal da forma da equação

$$A^T A = A A^T = 1 \quad (10)$$

cada transformação (uma rotação jacobi) é apenas uma rotação plana projetada para aniquilar um dos elementos da matriz fora da diagonal. A transformação sucessiva até os elementos fora da diagonal ficarem cada vez menores, até a matriz diagonal ter a precisão desejada. A acumulação de produtos de transformações fornece a matriz de autovetores, enquanto os elementos da matriz diagonal final são os valores próprios (WILLIAM, 2007).

#### 2.1.4. Etapa 5

O autovetor com o maior autovalor associado, corresponde à componente principal do conjunto de dados usados, essa componente é a mais significativa na dimensão dos dados. Através dela os autovetores devem ser ordenados na ordem de maior significância (de maior autovalor).

Aluno		RA/Matrícula	Professor	Tipo	
Cristiano Lopes Moreira		119103-0	Dr Reinaldo Bianchi	Relatório de implementação	
Data	Versão	Turma	Nome do arquivo		Página
10/11/2019	1	2º. Semestre de 2019	PEL_208_Relatório_02_Cristiano_Moreira.doc		6 (21)

Este processo irá possibilitar analisar a componente mais importante, como também filtrar as componentes de baixa relevância.

### 2.1.5. Novo conjunto de dados

Concluído o método PCA, basta recompor os dados pela equação

$$DadoOriginal = (Feature\ Vector^T \times DadoFinal) + MédiasOriginais \quad (11)$$

## 3. Proposta de implementação

### 3.1. Algoritmo de Análise de Componentes Principais PCA:

O algoritmo para realizar a análise de componentes principais irá utilizar, em sua base, as regras e operações matemáticas de matrizes, para tal será desenvolvido as rotinas: `matriz_DataAdjust`, rotina para formação da matriz de entrada normalizada pelas médias, recebe uma matriz A e retorna a matriz DataAdjust com a diferença dos pontos para a media; `matriz_covariancia`, rotina para formação da matriz de covariância, recebe uma matriz DataAdjust e retorna a matriz covariância  $\Sigma$ ; `matriz_autovalores`, rotina para cálculo dos autovalores - eigenvalues, recebe uma matriz covariância  $\Sigma$  e retorna a matriz de autovalores  $\Lambda$ ; `matriz_autovetores`, rotina para cálculo dos autovetores - eigenvectors, recebe uma matriz covariância  $\Sigma$  e a matriz de autovalores  $\Lambda$  e retorna a matriz de autovetores  $\Phi$ ; e `matriz_feature_vector`, rotina para cálculo dos Vetores principais - Feature Vector, recebe os auto valores e auto vetores ( $\Phi$ ,  $\Lambda$ ) e ordena os vetores de acordo com os eixos de maior importância

### 3.2. Pseudocódigos

**Matriz\_DataAdjust(matrizA)**  
 recebe matrizes  $A_{(n, m)}$  com dados para análise  
 $Media_{(m)} \leftarrow$  calcula a média de cada coluna m da matrizA  
 $MatrizADJ_{(i, j)} \leftarrow A_{(i, j)} - Media_{(j)}$   
 retorna MatrizADJ

Aluno <b>Cristiano Lopes Moreira</b>		RA/Matrícula <b>119103-0</b>	Professor <b>Dr Reinaldo Bianchi</b>	Tipo <b>Relatório de implementação</b>	
Data <b>10/11/2019</b>	Versão <b>1</b>	Turma 2º. Semestre de 2019	Nome do arquivo <b>PEL_208_Relatório_02_Cristiano_Moreira.doc</b>		Página <b>7 (21)</b>

```

matriz_covariancia(MatrizADJ)
  recebe matrizADJ(n, m) já ajustada pelas médias
  define Cbase(i, j) com 1 coluna e número de linhas igual a (n * m) de A
  loop para cada linha (n) da matrizADJ(n, m)
    loop para cada coluna (m) da matrizADJ(n, m)
      loop para cada coluna (k) da matrizCbase(i, j)
        Cbase(coluna*m+k) ← matrizADJ(i, j) * matrizADJ(i, k)
  Σ ← Cbase/(linhas-1)
  retorna Σ

matriz_autovalores(Σ)
  calcula raízes Λ pelo determinante (C - Λ) = 0
  retorna Λ

matriz_autovetores(Σ, Λ)
  forme as equações lineares com os equivalentes e a covariância
  Φ ← define as direções do vetor e calcula com cada equivalente
  Retorna Φ

matriz_feature_vector(Φ, Λ)
  ordena Φ decrescente pelo índice de do autovalor Λ
  retorna Φ

```

## 4. Experimentação e Resultados

Para verificar o funcionamento do algoritmo de análise das componentes principais, foi realizada a implementação em Python confrontando os diferentes resultados com três bases de dados:

- Dados do Censo Americano, com o tamanho da população dos Estados Unidos entre 1900 e 2000;
- Base de dados com 17 medições de temperatura x pressão nos Alpes;
- Base de dados com 40 amostras do resultado dos alunos sendo de 2 variáveis explicativas (livros e palestras) e 1 relacionada (nota).

Aluno		RA/Matrícula	Professor	Tipo	
Cristiano Lopes Moreira		119103-0	Dr Reinaldo Bianchi	Relatório de implementação	
Data	Versão	Turma	Nome do arquivo		Página
10/11/2019	1	2º. Semestre de 2019	PEL_208_Relatório_02_Cristiano_Moreira.doc		8 (21)



Ambiente:

PyCharm 2019.2.2 (Professional Edition) Build#PY-192.6603.34

Python 3.7.5 (tags/v3.7.5:5c02a39a0b, Oct 15 2019, 01:31:54) on win32

## Bibliotecas:

matplotlib-3.1.1      (utilizado para plotagem de gráficos)

pandas-0.25.2      (suporte à plotagem de gráficos)

xlrd-1.2.0      (leitura de arquivos do Excel - base de dados)

## Base de Dados:

bianchiDbAula1.xlsx      (Base uni-variáveis)

bianchiDbAula1MultiVariada.xlsx      (base multivariáveis)

Aluno		RA/Matrícula	Professor	Tipo	
Cristiano Lopes Moreira		119103-0	Dr Reinaldo Bianchi	Relatório de implementação	
Data	Versão	Turma	Nome do arquivo		Página
10/11/2019	1	2º. Semestre de 2019	PEL_208_Relatório_02_Cristiano_Moreira.doc		9 (21)

#### 4.1. Censo Americano

Base de dados:

Censo dos EUA – Valores Mensurados	
Ano	População [Mil]
1900	76.00
1910	91.97
1920	105.71
1930	123.20
1940	131.67
1950	150.70
1960	179.32
1970	203.21
1980	226.51
1990	249.63
2000	281.42

Aluno		RA/Matrícula	Professor	Tipo	
Cristiano Lopes Moreira		119103-0	Dr Reinaldo Bianchi	Relatório de implementação	
Data	Versão	Turma	Nome do arquivo		Página
10/11/2019	1	2º. Semestre de 2019	PEL_208_Relatório_02_Cristiano_Moreira.doc		10 (21)

Resultados:

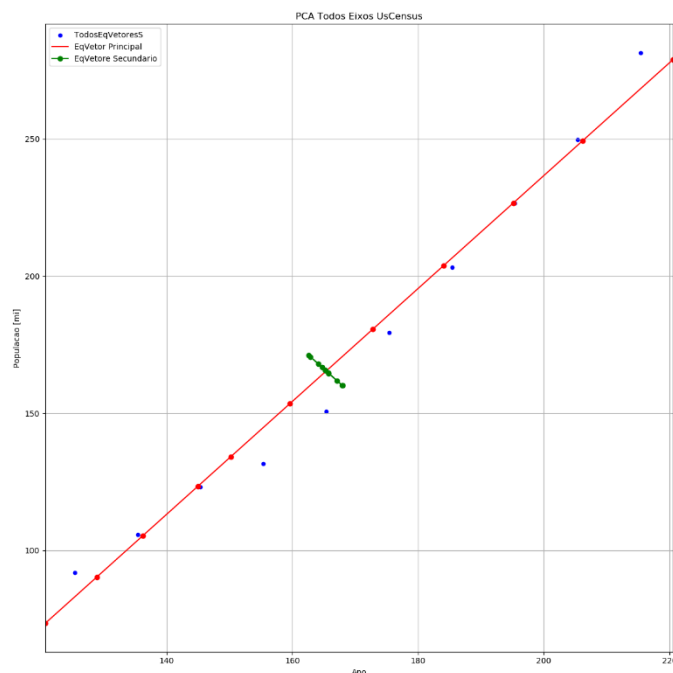


Fig2: Análise de comp. Principal USCensus [Ano x População]

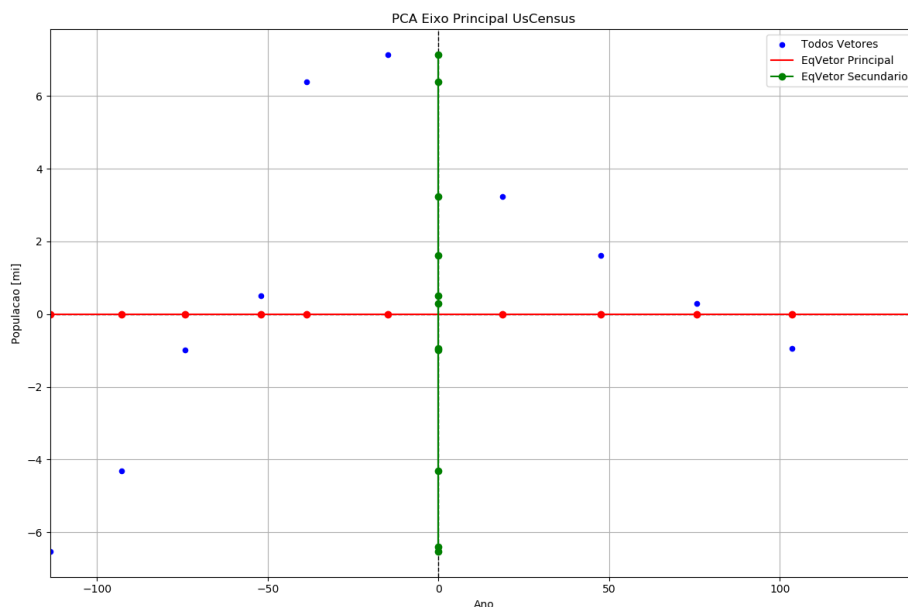


Fig3: Análise de comp. Principal USCensus [Ano x População] - eixo principal

Aluno		RA/Matrícula	Professor	Tipo	
Cristiano Lopes Moreira		119103-0	Dr Reinaldo Bianchi	Relatório de implementação	
Data	Versão	Turma	Nome do arquivo		Página
10/11/2019	1	2º. Semestre de 2019	PEL_208_Relatório_02_Cristiano_Moreira.doc		11 (21)

A representação do PCA nos gráficos das figuras 2 e 3 mostram a direção da componente principal vermelho, Ano, na relação com a população demográfica dos Estados Unidos. No eixo secundário verde, principalmente no gráfico da figura 2 em que é rotacionado sobre o eixo principal, podemos ver que a variação da componente secundária (população) não é linear, o que nos leva concluir da necessidade de uma verificação não linear desta componente.

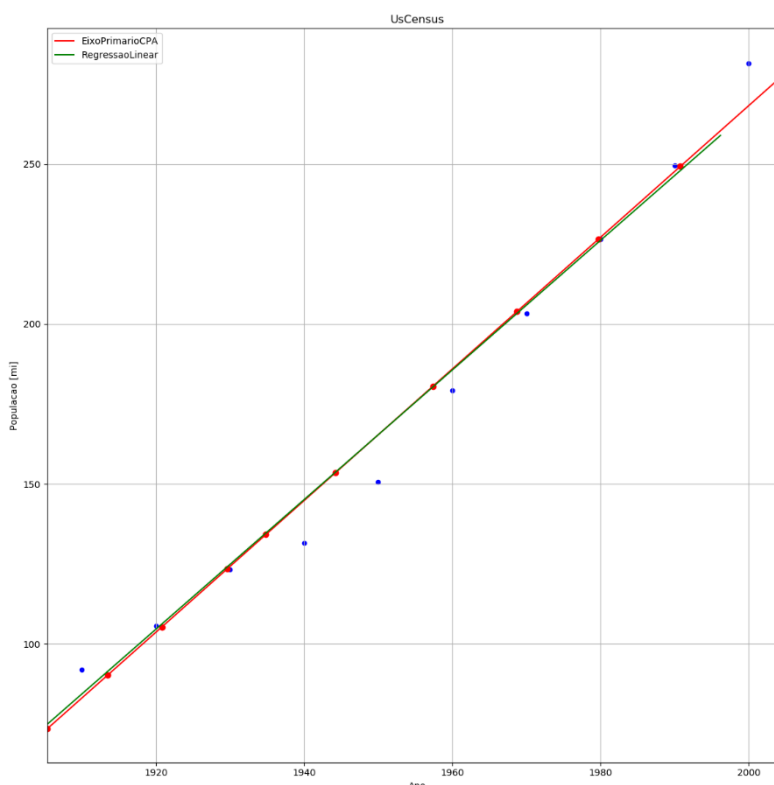


Fig4: Análise de comp. Principal USCensus [Ano x População] - LMS x PCA

A figura 4 mostra a relação entre a regressão linear pelo método dos mínimos quadrados e a análise pela componente principal, por ela é possível verificar um descolamento das duas projeções, sendo a PCA mais inclinada ao crescimento positivo da população e mais aderente com a realidade medida. Os modelos aparentam semelhança, mas tem significados distintos, o PCA é uma rotação do eixo de referência, e LMS um equilíbrio médio dos pontos sobre o eixo cartesiano (x,y).

Aluno		RA/Matrícula		Professor	Tipo
Cristiano Lopes Moreira		119103-0		Dr Reinaldo Bianchi	Relatório de implementação
Data	Versão	Turma		Nome do arquivo	Página
10/11/2019	1	2º. Semestre de 2019		PEL_208_Relatório_02_Cristiano_Moreira.doc	12 (21)

Censo dos EUA			
Regressão Linear			
Intercepto	$\hat{\beta}_0$	-81.064	
Coeficiente angular de Crescimento da População	$\hat{\beta}_1$	0.523	
PCA			
Covariância	$\Sigma$	$\begin{pmatrix} 1100 & 2227.83 \\ 2227.83 & 4599.59 \end{pmatrix}$	
Autovalores	$\Lambda$	$\begin{pmatrix} 16.949 & 5682.64 \end{pmatrix}$	
Autovetores	$\Phi$	$\begin{pmatrix} 1 & 0.4861 \\ -0.4861 & 1 \end{pmatrix}$	

Censo dos EUA – População [Mil] Componente Principal			
Ano	Linear Simples	Ano Ref CPA	Comp. Principal
1905	74.84	1905	73.42
1914	93.25	1913	90.27
1923	111.66	1921	105.32
1933	130.08	1930	123.40
1942	148.49	1935	134.18
1951	166.90	1944	153.50
1960	185.31	1957	180.59
1969	203.72	1969	203.84
1978	222.14	1980	226.62
1987	240.55	1991	249.25
1996	258.96	2005	278.90

Aluno		RA/Matrícula		Professor		Tipo	
Cristiano Lopes Moreira		119103-0		Dr Reinaldo Bianchi		Relatório de implementação	
Data	Versão	Turma		Nome do arquivo			Página
10/11/2019	1	2º. Semestre de 2019		PEL_208_Relatório_02_Cristiano_Moreira.doc			13 (21)

## 4.2. Pressão x Temperatura de Ebulição da água nos Alpes

Base de dados:

Ebulição da água – Valores Mensurados	
Temperatura [Fº]	Pressão [cmHg]
194.5	20.79
194.3	20.79
197.9	22.4
198.4	22.67
199.4	23.15
199.9	23.35
200.9	23.89
201.1	23.99
201.4	24.02
201.3	24.01
203.6	25.14
204.6	26.57
209.5	28.49
208.6	27.76
210.7	29.04
211.9	29.88
212.2	30.06

Aluno		RA/Matrícula	Professor	Tipo	
Cristiano Lopes Moreira		119103-0	Dr Reinaldo Bianchi	Relatório de implementação	
Data	Versão	Turma	Nome do arquivo		Página
10/11/2019	1	2º. Semestre de 2019	PEL_208_Relatório_02_Cristiano_Moreira.doc		14 (21)

Resultados:

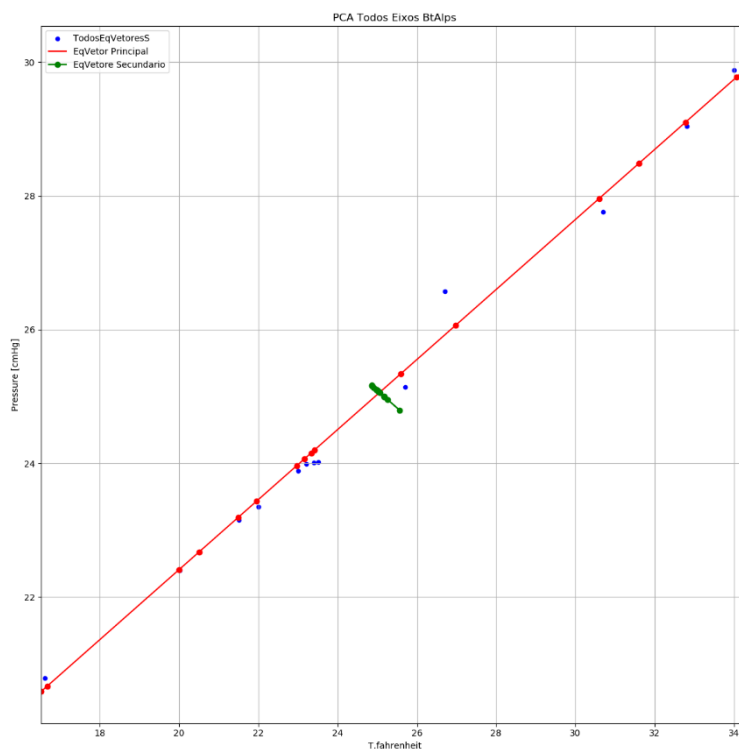


Fig5: Análise de comp. Principal Ebulição da água [t F° x Pressão]

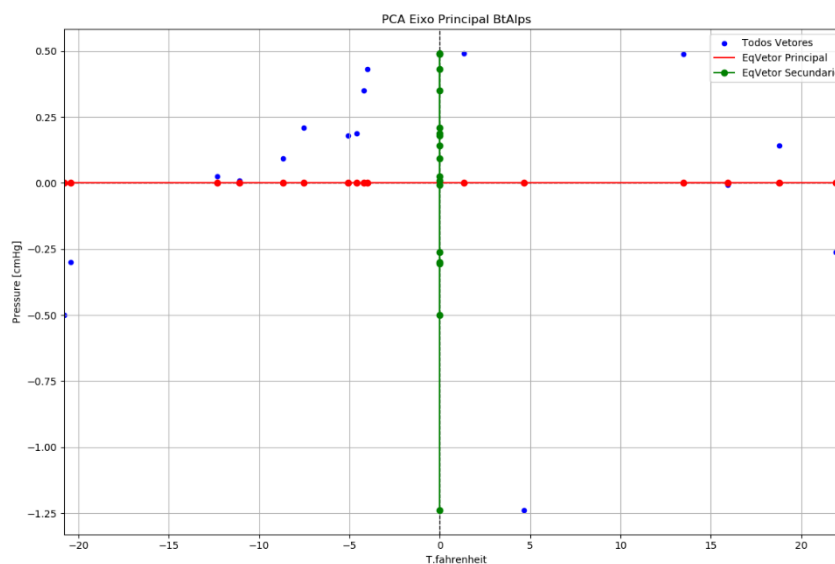


Fig6: Análise de comp. Principal Ebulição da água [t F° x Pressão] - eixo principal

Aluno		RA/Matrícula		Professor		Tipo	
Cristiano Lopes Moreira		119103-0		Dr Reinaldo Bianchi		Relatório de implementação	
Data	Versão	Turma		Nome do arquivo			Página
10/11/2019	1	2º. Semestre de 2019		PEL_208_Relatório_02_Cristiano_Moreira.doc			15 (21)

A representação do PCA nos gráficos das figuras 5 e 6 mostram a direção da componente principal vermelho, Temperatura de ebulição da água em [F°], na relação com a pressão atmosférica. No eixo secundário verde, podemos verificar que a variação da componente secundária é muito pequena e quase linear.

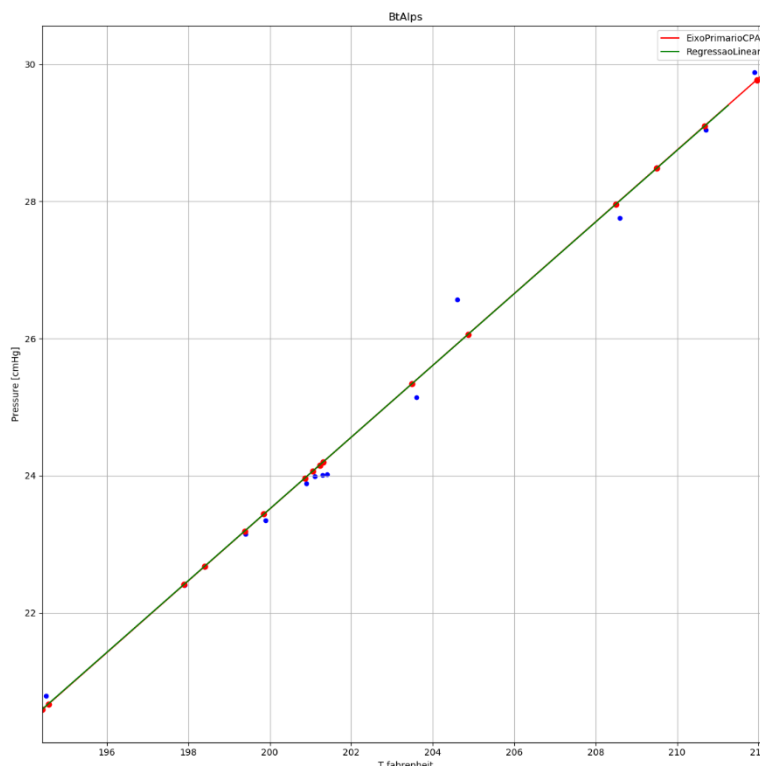


Fig7: Análise de comp. Principal Ebulição da água [t F° x Pressão] - LMS x PCA

A figura 7 mostra a relação entre a regressão linear pelo método dos mínimos quadrados e a análise pela componente principal, por ela é possível verificar que ambas as curvas, LMS e PCA, tem quase a mesma inclinação, uma característica de modelos bem lineares.

Aluno		RA/Matrícula		Professor	Tipo
Cristiano Lopes Moreira		119103-0		Dr Reinaldo Bianchi	Relatório de implementação
Data	Versão	Turma		Nome do arquivo	Página
10/11/2019	1	2º. Semestre de 2019		PEL_208_Relatório_02_Cristiano_Moreira.doc	16 (21)



Ebulição da água			
Regressão Linear			
Intercepto	$\hat{\beta}_0$	-81.064	
Coeficiente angular de T[F°]	$\hat{\beta}_1$	0.523	
PCA			
Covariância	$\Sigma$	$\begin{pmatrix} 33.17 & 17.34 \\ 17.34 & 9.12 \end{pmatrix}$	
Autovalores	$\Lambda$	$\begin{pmatrix} 0.0398 & 422.55 \end{pmatrix}$	
Autovetores	$\Phi$	$\begin{pmatrix} 1 & 1.9101 \\ -1.9101 & 1 \end{pmatrix}$	

Ebulição da água – Pressão [cmHg] - Componente Principal			
Temperatura [F°]	Linear Simples	Temperatura [F°]	Quadrática
194.41	20.59	194.41	20.59
195.46	21.14	194.56	20.67
196.51	21.69	197.89	22.41
197.57	22.24	198.40	22.67
198.62	22.79	199.38	23.19
199.67	23.34	199.85	23.44
200.72	23.89	200.86	23.96
201.78	24.44	201.06	24.07
202.83	24.99	201.22	24.15
203.88	25.55	201.31	24.20
204.94	26.10	203.49	25.34
205.99	26.65	204.87	26.06
207.04	27.20	208.50	27.96
208.10	27.75	209.50	28.49
209.15	28.30	210.67	29.10
210.20	28.85	211.96	29.77
211.25	29.40	212.27	29.93

Aluno		RA/Matrícula		Professor		Tipo	
Cristiano Lopes Moreira		119103-0		Dr Reinaldo Bianchi		Relatório de implementação	
Data	Versão	Turma		Nome do arquivo			Página
10/11/2019	1	2º. Semestre de 2019		PEL_208_Relatório_02_Cristiano_Moreira.doc			17 (21)

### 4.3. Livros, palestras e nota

Base de dados:

Livros x Palestras x Notas		
Nota	Livros	Palestras
45	0	9
57	1	15
45	0	10
51	2	16
65	4	10
88	4	20
44	1	11
87	4	20
89	3	15
59	0	15
66	2	8
65	1	13
56	4	18
47	1	10
66	0	8
41	1	10
56	3	16
37	0	11
45	1	19
58	4	12
47	4	11
64	0	19
97	2	15
55	3	15
51	1	20
61	0	6
69	3	15
79	3	19
71	2	14
62	2	13
87	3	17
43	2	11

Aluno		RA/Matrícula	Professor	Tipo	
Cristiano Lopes Moreira		119103-0	Dr Reinaldo Bianchi	Relatório de implementação	
Data	Versão	Turma	Nome do arquivo		Página
10/11/2019	1	2º. Semestre de 2019	PEL_208_Relatório_02_Cristiano_Moreira.doc		18 (21)

Resultados:

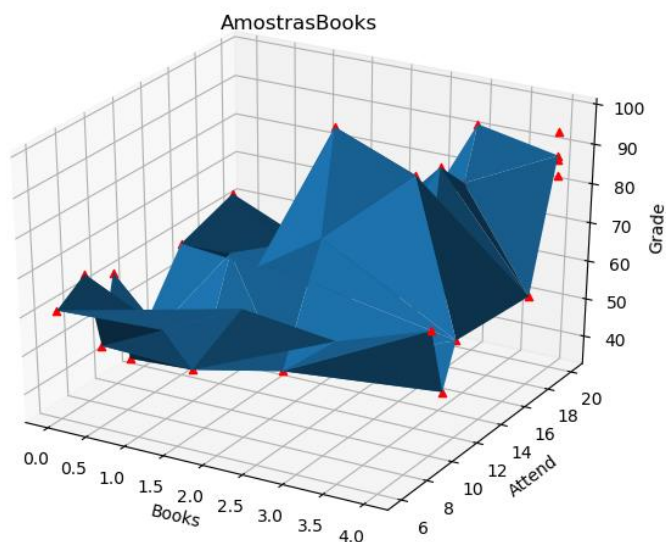


Fig8: Análise de comp. Principal Livros x Palestras e Notas

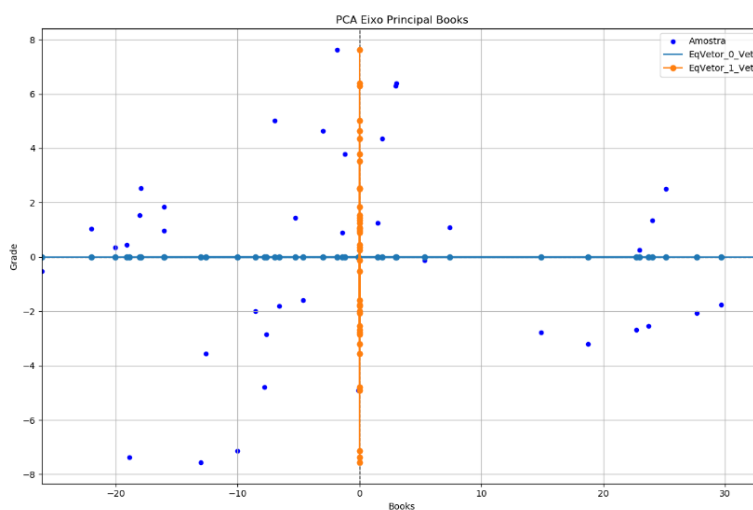


Fig9: Análise de comp. Principal Livros x Palestras e Aulas - eixo principal – Redução do Eixo Palestras

A fig8 mostra apenas a representação gráfica das amostras da eficiência dos alunos pela relação de leitura de livros e presença em seminários. Pelo método da análise dos componentes principais (PCA) foi realizada a redução da componente que menos descreve os dados, a presença em seminários, e foi possível criar o gráfico da Figura 9, que mostra o mapeamento da relação de leitura de livros e a nota do aluno.

Aluno		RA/Matrícula		Professor	Tipo
<b>Cristiano Lopes Moreira</b>		<b>119103-0</b>		<b>Dr Reinaldo Bianchi</b>	<b>Relatório de implementação</b>
Data	Versão	Turma		Nome do arquivo	Página
<b>10/11/2019</b>	<b>1</b>	<b>2º. Semestre de 2019</b>		<b>PEL_208_Relatório_02_Cristiano_Moreira.doc</b>	<b>19 (21)</b>

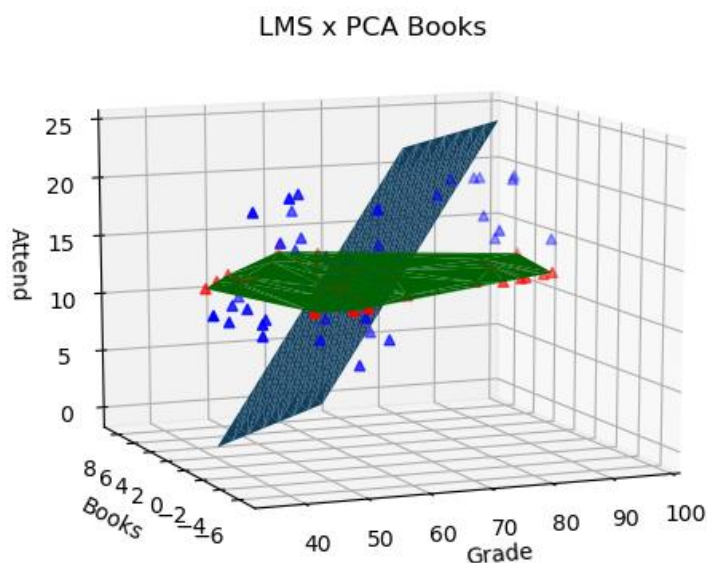


Fig10: Análise de comp. Principal Livros x Palestras e Notas - LMS x PCA (redução do eixo palestras)

O PCA e a regressão dos linear das evidências de leitura e participação de palestras pela nota do aluno mostra que, pelos autovalores (PCA) e pelo coeficiente angular da variável (LMS) explicativa livros, que a leitura traz 2,5 vezes mais resultados nas notas dos alunos que a participação em palestras. E aparenta ser, pela redução da componente menos significativa, possível descrever o modelo somente pela variável de leitura de livros.

Eficiência de alunos			
Regressão Linear			
Intercepto	$\hat{\beta}_0$		34.942
Coeficiente angular de Livros	$\hat{\beta}_1$		3.501
Coeficiente angular de Palestras	$\hat{\beta}_2$		1.364
PCA			
Covariância	$\Sigma$	$\begin{pmatrix} 279.07 & 11.77 & 34.45 \\ 11.76 & 2.05 & 2.71 \\ 34.45 & 2.717 & 18.29 \end{pmatrix}$	
Autovalores	$\Lambda$	$(284.06 \quad 13.93 \quad 1.42)$	
Autovetores	$\Phi$	$\begin{pmatrix} 0.99 & 0.132 & 0.029 \\ 0.0425 & -0.094 & -0.99 \\ 0.128 & -0.986 & 0.099 \end{pmatrix}$	

Aluno		RA/Matrícula	Professor	Tipo	
<b>Cristiano Lopes Moreira</b>		<b>119103-0</b>	<b>Dr Reinaldo Bianchi</b>	<b>Relatório de implementação</b>	
Data	Versão	Turma	Nome do arquivo		Página
<b>10/11/2019</b>	<b>1</b>	2º. Semestre de 2019	<b>PEL_208_Relatório_02_Cristiano_Moreira.doc</b>		<b>20 (21)</b>

## 5. Conclusão

O método da análise dos componentes principais (PCA) oferece uma oportunidade para a criação de modelos estatísticos com a segmentação de elementos de uma massa de dados difusa, na ótica de eixos/componentes de maior relevância, de forma a auxiliar a explicação de cada parâmetro (variável explicativa) sobre o fenômeno dependente (variável dependente). Possibilita também a redução e simplificação do modelo e da amostra pela exclusão de componentes de baixa relevância.

## 6. Referências

- [1] FÁVERO, Luiz Paulo et al. **Análise de dados**: modelagem multivariada para tomada de decisões. Rio de Janeiro: Elsevier, 2009.
- [2] HASTIE, Trevor; TIBSHIRANI, Robert; FRIEDMAN, Jerome. **The Elements of Statistical Learning**: Data Mining, Inference, and Prediction. 2. ed. Stanford: Springer, 2008.
- [3] WILBUR, Anderson Theodore. **An introduction to Multivariate Statistical Analysis**. 2. ed. Stanford: Wiley, 1971.
- [4] PRESS, William H. et al. **Numerical Recipes**: The Art of Scientific Computing. 3. ed. Cambridge, Massachusetts: Cambridge University Press, 2007.

Aluno		RA/Matrícula		Professor		Tipo	
Cristiano Lopes Moreira		119103-0		Dr Reinaldo Bianchi		Relatório de implementação	
Data	Versão	Turma		Nome do arquivo			Página
10/11/2019	1	2º. Semestre de 2019		PEL_208_Relatório_02_Cristiano_Moreira.doc			21 (21)