

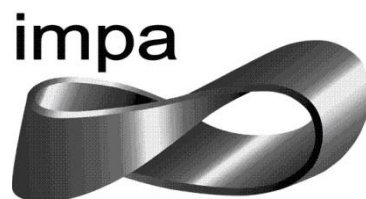
INSTITUTO NACIONAL DE MATEMÁTICA PURA E APLICADA
MESTRADO PROFISSIONAL EM MATEMÁTICA

Rômulo de Macedo Ciraudó

**O uso da Estatística como ferramenta de análise de
resultado de avaliação**

Rio de Janeiro

Fevereiro de 2015



INSTITUTO NACIONAL DE MATEMÁTICA PURA E APLICADA
MESTRADO PROFISSIONAL EM MATEMÁTICA

Rômulo de Macedo Ciraudó

O uso da Estatística como ferramenta de análise de resultado de avaliação

Trabalho de Conclusão de Curso do Mestrado Profissional em Matemática em Rede Nacional, apresentado ao Instituto Nacional de Matemática Pura e Aplicada como requisito final para a obtenção do título de Mestre.

Orientador: Professor Mestre Eduardo Wagner.

Rio de Janeiro

Fevereiro de 2015

INSTITUTO NACIONAL DE MATEMÁTICA PURA E APLICADA
MESTRADO PROFISSIONAL EM MATEMÁTICA

Rômulo de Macedo Ciraudó

**O uso da Estatística como ferramenta de análise de
resultado de avaliação**

Trabalho de Conclusão de Curso do Mestrado Profissional em Matemática em Rede Nacional, apresentado ao Instituto Nacional de Matemática Pura e Aplicada como requisito final para a obtenção do título de Mestre.

Data de aprovação: ____/____/____

Banca Examinadora:

PROF. EDUARDO WAGNER - FGV (ORIENTADOR)

PROF. PAULO CEZAR PINTO CARVALHO - IMPA / FGV

PROF. ANTONIO CARLOS SARAIVA BRANCO - FGV

Agradecimentos

Agradeço a todos que contribuíram, direta ou indiretamente, para que este trabalho fosse realizado. E em particular:

a Deus, por me guiar e proteger em todos os momentos;

aos meus pais pelo amor e pela minha formação pessoal;

à minha esposa e minhas filhas, por todo amor, pelo incentivo e por serem os principais motivos que me levam a ser uma pessoa melhor a cada dia;

a todos os Professores do IMPA que ministraram aulas durante o curso, em especial ao Prof. Eduardo Wagner pelo acolhimento, pelo apoio e pela orientação para o desenvolvimento deste trabalho;

aos amigos pela presença e pelo apoio em todas as horas, principalmente nas mais difíceis,

a CAPES,

ao IMPA.

Resumo

Este trabalho sugere uma reflexão acerca dos métodos utilizados pelos profissionais da educação para analisar os resultados das avaliações às quais os alunos são submetidos, e tem por objetivo motivar esses profissionais - em especial os professores e os coordenadores pedagógicos – a se apropriarem e utilizarem algumas ferramentas estatísticas a fim de auxiliar na interpretação de dados e conseqüentemente nas tomadas de decisão que afetam, direta ou indiretamente, o processo ensino-aprendizagem. Apresentamos de forma simples e direta alguns conceitos importantes e bastante úteis para os objetivos propostos. Exemplificamos como utilizar uma poderosa ferramenta estatística capaz de nortear tais profissionais no sentido de identificar a homogeneidade do aprendizado. Comparamos resultados provenientes de diferentes avaliações. Buscamos respostas para algumas perguntas do tipo: Os alunos que tiram boas notas em Matemática, também tiram boas notas em Física? Esperamos, verdadeiramente, que ao final da leitura desse trabalho, o leitor esteja convencido de que a utilização das ferramentas estatísticas aqui sugeridas auxiliá-lo-á a encontrar algumas respostas para questionamentos que permeiam o cotidiano escolar.

Palavras-chave: estatística, avaliação, desvio-padrão, escore padronizado, correlação.

Abstract

This work suggests a reflection on the methods used by education professionals to analyze the results of evaluations to which students are subjected, and aims to motivate these professionals - particularly teachers and coordinators - to appropriate and use some tools statistics to assist in the interpretation of data and consequently in making decisions that affect, directly or indirectly, the teaching-learning process. Here is a simple and direct way some important concepts and quite useful for those goals. Exemplify using a powerful statistical tool able to guide such professionals to identify the homogeneity of learning. We compare results from different evaluations. We seek answers to some questions like: Students who take good grades in Mathematics, also take good grades in physics? We hope, indeed, that the end of the reading of this work, the reader is convinced that the use of statistical tools suggested here will help you find some answers to questions that permeate the school routine.

Keywords: statistics, evaluation, standard deviation, standard score, correlation.

Sumário

1. Introdução.....	8
2. Os conceitos.....	11
2.1. Estatística.....	12
2.1.1. População e Amostra.....	12
2.2. Frequências.....	13
2.2.1. Frequência absoluta.....	13
2.2.2. Frequência relativa.....	14
2.2.3. Frequência acumulada.....	15
2.2.4. Frequência acumulada relativa.....	17
2.3. Medidas de tendência central.....	18
2.3.1. Média aritmética.....	18
2.3.2. Mediana.....	20
2.3.3. Moda.....	21
2.4. Medidas de Dispersão.....	22
2.4.1. Variância.....	22
2.4.2. Desvio padrão.....	23
2.5. Escore padronizado.....	28
2.6. Correlação linear.....	30
3. A utilização do desvio-padrão.....	37
3.1. Regra Empírica ou Regra 68-95-99.....	41
3.2. Teorema de Tchebichev.....	41

4. Escore padronizado, como utilizar?.....	43
5. Existe relação entre as notas das disciplinas?.....	49
6. Conclusão.....	64
7. Referências bibliográficas.....	66

Capítulo 1

Introdução

A Estatística pode ser vista em suas inúmeras aplicações como um método de descobrir as relações de dependências entre causas e efeitos de um fenômeno qualquer, um instrumento lógico, fundado em um método indutivo, que visa descobertas para a melhoria do universo estudado.

Em geral, as informações estatísticas são eficazes e específicas, fornecendo assim subsídios importantíssimos para as tomadas de decisões racionais. Neste sentido, a estatística fornece ferramentas úteis para que as instituições possam definir melhor suas metas, avaliar seu desempenho, identificar seus pontos fracos e atuar na melhoria contínua de seus processos.

Um dos objetivos contemplados pelo presente trabalho é, justamente, nortear os profissionais da educação no sentido de que possam, através de ferramentas estatísticas, interpretar dados e transformá-los em informações importantes que muitas das vezes estão ocultas. Para caminharmos a encontro dos nossos objetivos, no decorrer do capítulo 2, apresentaremos alguns conceitos comumente estudados na estatística, que facilitarão e auxiliarão no desenvolvimento dos assuntos dos capítulos subsequentes.

No que tange o assunto avaliação, é muito comum, ao se deparar com as notas obtidas pelos alunos de uma turma, a maioria dos educadores utilizarem como parâmetro de avaliação do resultado apenas a média das notas.

Uma breve reflexão sobre as medidas de tendência central permite-nos concluir que elas não são suficientes para caracterizar totalmente uma sequência numérica.

Se observarmos as sequências:

X: 10, 1, 18, 20, 35, 3, 7, 15, 11, 10.

Y: 12, 13, 13, 14, 12, 14, 12, 14, 13, 13.

Z: 13, 13, 13, 13, 13, 13, 13, 13, 13, 13.

concluiremos que todas possuem a mesma média 13.

No entanto, são sequências completamente distintas do ponto de vista de variabilidade de dados.

Na sequência Z não há variabilidade de dados.

A média 13 representa bem qualquer valor da série.

Na sequência Y, a média 13 representa bem a série, mas existem elementos da série levementes diferenciados da média 13.

Na sequência X, existem muitos elementos bastante diferenciados da média 13.

Concluimos que a média 13 representa otimamente a sequência Z, representa bem a sequência Y, mas não representa bem a sequência X. SILVA et al. (1996, p.100)

Convencido de que a utilização de apenas uma medida de tendência central não é suficiente para conhecermos bem as características de um conjunto de dados, no capítulo 3 sugerimos a utilização de uma medida de dispersão para que se tenha uma análise mais completa.

No primeiro capítulo de um de seus livros, D'HAINAUT (1977) aborda o seguinte questionamento: "Mandemos agora fazer duas provas de matemática a duas turmas (A e B). Em cada turma, os resultados dos alunos são muito variáveis (...). No conjunto, pode parecer-nos que os resultados da turma A são superiores aos da turma B, mas isso não impede que muitos dos alunos de A não tenham resultados inferiores, relativamente a determinados alunos da turma B". Retornaremos nesse assunto no capítulo 4, onde abordaremos a utilização do escore padrão.

Por se tratar de um universo de experiências humanas e diversificadas, a educação apresenta-se como um campo científico repleto de sutilezas e ambiguidades. Nesta perspectiva, devemos estar atentos as peculiaridades de cada situação, lançando mão das diversas ferramentas disponíveis para um diagnóstico preciso e conseqüentemente a adoção das medidas adequadas.

No capítulo 5, através de uma pequena amostragem, fizemos uma breve análise sobre a existência ou não de relação entre as notas obtidas pelos alunos em determinadas disciplinas ministradas no ensino médio.

Definitivamente, não faz parte dos objetivos desse trabalho afirmar que os critérios e métodos apresentados são absolutos e insubstituíveis, nem tão pouco esgotar os temas apresentados. O que pretendemos é apontar caminhos que acreditamos nos levar na direção de conclusões coerentes.

Capítulo 2

Os conceitos

Para facilitar o entendimento de cada conceito apresentado neste capítulo, adotaremos notas obtidas aleatoriamente de uma turma fictícia e exemplificaremos cada conceito. Este processo fora adotado para que os resultados fossem mais simples e conseqüentemente mais adequados do ponto de vista didático.

Abaixo temos as notas dos alunos:

Número do aluno	Nota do aluno
1	5,0
2	8,0
3	9,0
4	2,0
5	7,0
6	8,0
7	9,0
8	7,0
9	5,0
10	5,0
11	9,0
12	10,0
13	8,0
14	6,0
15	10,0

Número do aluno	Nota do aluno
16	7,0
17	8,0
18	6,0
19	4,0
20	0,0
21	3,0
22	1,0
23	6,0
24	2,0
25	5,0
26	7,0
27	9,0
28	8,0
29	6,0
30	5,0

2.1. Estatística

Antes mesmo do surgimento da palavra “Estatística”, historicamente temos que:

Desde a Antiguidade, vários povos já registravam o número de habitantes, de nascimentos, de óbitos, e faziam estimativas das riquezas individuais sociais, distribuíam equitativamente terras ao povo, cobravam impostos e realizavam inquéritos quantitativos por processos que, hoje chamaríamos de “estatística”. (Leitão, 2011, p.18)

As várias facetas da estatística nos levam a dificuldade de conceituá-la de forma satisfatoriamente precisa. Nesse sentido, apresentaremos alguns dos conceitos encontrados ao longo da pesquisa sob a ótica dos respectivos autores.

Para Triola (1999, p.2), “a estatística é uma coleção de métodos para planejar experimentos, obter dados e organizá-los, resumi-los, analisá-los, interpretá-los e deles extrair conclusões”.

De acordo com Milone (2006, p.3), “Estatística é o estudo dos modos de obtenção, coleta, organização, processamento e análise de informações relevantes que permitem quantificar, qualificar ou ordenar entes, coleções, fenômenos ou populações de modo tal que se possa concluir, deduzir ou prever propriedades, eventos ou estados futuros”.

2.1.1. População e Amostra

Podemos afirmar que uma população “é um conjunto de indivíduos ou objetos que apresentam pelo menos uma característica em comum”. (FONSECA e MARTINS, 2012, p.111)

De acordo com Triola (1999), uma amostra é um subconjunto de membros selecionados de uma população.

Tendo em vista as definições acima, fica estabelecido que os dados utilizados neste trabalho são dados amostrais e não devem ser confundidos com dados populacionais.

2.2. Frequências

2.2.1. Frequência Absoluta

Segundo COSTA NETO (1977, p.8), “O primeiro passo para se descrever graficamente um conjunto de dados observados é verificar as *frequências* dos diversos valores existentes da variável. Definimos a frequência de um dado valor de uma variável (qualitativa ou quantitativa) como o número de vezes que esse valor foi observado”. Chamaremos, de agora em diante, a frequência descrita pela definição acima como “Frequencia Absoluta” e para ela utilizaremos a notação “ f_i ”.

Exemplo:

Abaixo temos a distribuição de frequência das notas dos alunos da turma utilizada para exemplo descrita no início desse capítulo.

Notas	Frequência absoluta (f_i)
0,0	1
1,0	1
2,0	2
3,0	1
4,0	1
5,0	5
6,0	4
7,0	4
8,0	5
9,0	4

10,0	2
Somatório (Σ)	30

Ao observarmos a tabela de distribuição de frequências, torna-se bastante fácil a seguinte interpretação: um aluno obteve nota zero, um aluno obteve nota um, dois alunos obtiveram nota dois, um aluno obteve nota três, um aluno obteve nota quatro, cinco alunos obtiveram nota cinco, quatro alunos obtiveram nota seis, quatro alunos obtiveram nota sete, cinco alunos obtiveram nota oito, quatro alunos obtiveram nota nove e dois alunos obtiveram nota dez.

Uma vez que os dados estejam colocados na forma de uma distribuição de frequências, com as respectivas frequências absolutas, poderemos de forma simples obter informações adicionais e bastante úteis se considerarmos os seguintes conceitos:

2.2.2. Frequência Relativa (f_r)

A Frequência Relativa de um elemento é dada pela razão entre a frequência absoluta deste elemento e o número total de elementos da distribuição (somatório das frequências absolutas). Esta frequência pode ser expressa na forma de razão, na forma decimal ou na forma de porcentagem.

Em sua forma de porcentagem, torna-se de fácil interpretação quanto à proporção que o elemento aparece em relação ao total de elementos da distribuição.

Exemplo:

Notas	Frequência absoluta (f_i)	Frequência relativa (f_r)
0,0	1	1/30 ou 0,0333... ou 3,33%
1,0	1	1/30 ou 0,0333... ou 3,33%
2,0	2	2/30 ou 0,0666... ou 6,67%

3,0	1	1/30 ou 0,0333... ou 3,33%
4,0	1	1/30 ou 0,0333... ou 3,33%
5,0	5	5/30 ou 0,1666... ou 16,67%
6,0	4	4/30 ou 0,1333... ou 13,33%
7,0	4	4/30 ou 0,1333... ou 13,33%
8,0	5	5/30 ou 0,1666... ou 16,67%
9,0	4	4/30 ou 0,1333... ou 13,33%
10,0	2	2/30 ou 0,0666... ou 6,67%
Somatório (Σ)	30	30/30 ou 1,0 ou 100%

Observações:

- Os percentuais que aparecem na tabela acima sofreram arredondamento e por isso seu somatório resulta em 99,99%, todavia, para efeito didático arredondamos o somatório para 100%.

- Algumas das frações apresentadas na tabela acima não estão em sua forma irredutível. Isso se deve ao fato de não ser conveniente que se faça a simplificação, pois a utilização de frações irredutíveis pode tornar a interpretação menos trivial.

2.2.3. Frequência Acumulada (F_i)

Para obtermos a frequência acumulada de um elemento, devemos somar a frequência absoluta deste elemento com as frequências absolutas dos elementos que o antecedem na tabela de distribuição de frequências.

Em geral, a frequência acumulada é muito útil quando se deseja obter a quantidade de dados que estão abaixo ou acima de um valor determinado. Podemos utilizar essa frequência,

por exemplo, para determinar quantos alunos obtiveram nota abaixo da média em uma determinada avaliação.

Exemplo:

Notas	Frequência absoluta (f_i)	Frequência acumulada (F_i)
0,0	1	1
1,0	1	2
2,0	2	4
3,0	1	5
4,0	1	6
5,0	5	11
6,0	4	15
7,0	4	19
8,0	5	24
9,0	4	28
10,0	2	30
Somatório (Σ)	30	-----

Observações:

- Observando a tabela acima, fica bastante fácil identificar, por exemplo, que dezenove alunos obtiveram notas menores ou iguais a sete. E conseqüentemente, pela diferença, onze alunos obtiveram notas maiores que sete.

- Não se faz necessário o cálculo do somatório das frequências acumuladas, pois esse resultado não possui uma interpretação útil ao nosso estudo.

2.2.4. Frequência Acumulada Relativa (F_{ri})

É a divisão da frequência acumulada do elemento pelo somatório das frequências absolutas. Sua utilidade é semelhante a da frequência relativa, porém deve ser expressa na forma de razão, na forma decimal ou na forma de porcentagem.

Notas	Frequência absoluta (f_i)	Frequência acumulada (F_i)	Frequência acumulada relativa (F_{ri})
0,0	1	1	1/30 ou 0,0333... ou 3,33%
1,0	1	2	2/30 ou 0,0666... ou 6,67%
2,0	2	4	4/30 ou 0,1333... ou 13,33%
3,0	1	5	5/30 ou 0,1666... ou 16,67%
4,0	1	6	6/30 ou 0,2... ou 20%
5,0	5	11	11/30 ou 0,3666... ou 36,67%
6,0	4	15	15/30 ou 0,5... ou 50%
7,0	4	19	19/30 ou 0,6333... ou 63,33%
8,0	5	24	24/30 ou 0,8... ou 80%
9,0	4	28	28/30 ou 0,9333... ou 93,33%
10,0	2	30	30/30 ou 1,0... ou 100%
Somatório (Σ)	30	-----	-----

Observações:

- Observando a tabela acima, fica bastante fácil identificar, por exemplo, que oitenta por cento dos alunos obtiveram notas menores ou iguais a oito. E conseqüentemente, pela diferença, vinte por cento dos alunos obtiveram notas maiores que oito.

- Não se faz necessário o cálculo do somatório das frequências relativas acumuladas, pois esse resultado não possui uma interpretação útil ao nosso estudo.

2.3. Medidas de tendência central

TRIOLA (1999), em seu livro “Introdução à Estatística”, nos contempla com as seguintes definições:

“A média aritmética de um conjunto de valores é a medida de centro encontrada pela adição dos valores e divisão do total pelo número de valores.

A mediana de um conjunto de dados é a medida de centro que é o valor do meio quando os dados originais estão arranjados em ordem crescente (ou decrescente) de magnitude.

A moda de um conjunto de dados é o valor que ocorre mais frequentemente.”

Utilizaremos a mesma distribuição de frequência das notas para exemplificar e cada um dos conceitos acima.

2.3.1 Média

Calcular a média aritmética de um conjunto de valores é uma tarefa bastante simples. Para isso, basta somar os valores de todas as notas e efetuar a divisão pela quantidade de notas.

De acordo com Fonseca (2012), quando os valores estiverem agrupados em uma tabela de distribuição de frequências, para facilitar o cálculo da média utilizaremos a idéia da média aritmética ponderada. Para isso, multiplicaremos cada nota pela sua respectiva frequência absoluta (peso) e somaremos os produtos obtidos. Após encontrar o resultado dessa soma, dividiremos o resultado pela quantidade de notas.

Observe que os dois processos são equivalentes entre si e se diferem apenas pelo agrupamento das notas iguais em uma mesma classe de frequência e como consequência a utilização da multiplicação (adição de parcelas de mesmo valor).

Notas	Frequência absoluta (f_i)	Nota x Frequência
0,0	1	$0,0 \cdot 1 = 0,0$
1,0	1	$1,0 \cdot 1 = 1,0$
2,0	2	$2,0 \cdot 2 = 4,0$
3,0	1	$3,0 \cdot 1 = 3,0$
4,0	1	$4,0 \cdot 1 = 4,0$
5,0	5	$5,0 \cdot 5 = 25,0$
6,0	4	$6,0 \cdot 4 = 24,0$
7,0	4	$7,0 \cdot 4 = 28,0$
8,0	5	$8,0 \cdot 5 = 40,0$
9,0	4	$9,0 \cdot 4 = 36,0$
10,0	2	$10,0 \cdot 2 = 20,0$
Somatório (Σ)	30	185,0

Para finalmente encontrarmos o valor da média, devemos efetuar a divisão de 185,0 por 30 e concluirmos que a média das notas é igual a 6,17

Conforme afirma MILONE (2006), a média de um conjunto de dados possui as seguintes características:

- Sempre existe;
- É única;
- É afetada por todos os valores da distribuição. Isso faz com que seu valor se altere sempre que se alterar um valor da distribuição.

2.3.2. Mediana

Para determinar o valor da mediana, utilizaremos a tabela de distribuição de frequências após determinarmos os valores da coluna da frequência acumulada relativa. A mediana será a nota que possuir a menor frequência relativa acumulada que seja maior que cinquenta por cento. Caso exista uma frequência relativa acumulada exatamente igual à cinquenta por cento, a mediana será igual à média aritmética da nota que possuir tal frequência e a nota da classe seguinte que possua frequência absoluta diferente de zero.

Notas	Frequência absoluta (f_i)	Frequência acumulada (F_i)	Frequência acumulada relativa (F_{ri})
0,0	1	1	1/30 ou 0,0333... ou 3,33%
1,0	1	2	2/30 ou 0,0666... ou 6,67%
2,0	2	4	4/30 ou 0,1333... ou 13,33%
3,0	1	5	5/30 ou 0,1666... ou 16,67%
4,0	1	6	6/30 ou 0,2... ou 20%
5,0	5	11	11/30 ou 0,3666... ou 36,67%
6,0	4	15	15/30 ou 0,5... ou 50%
7,0	4	19	19/30 ou 0,6333... ou 63,33%
8,0	5	24	24/30 ou 0,8... ou 80%

9,0	4	28	28/30 ou 0,9333... ou 93,33%
10,0	2	30	30/30 ou 1,0... ou 100%
Somatório (Σ)	30	-----	-----

Para o exemplo acima, temos que a nota seis possui frequência acumulada relativa exatamente igual à cinquenta por cento, dessa forma a mediana será a média aritmética entre as notas seis e sete.

$$\text{Mediana} = \frac{6,0 + 7,0}{2} = 6,5$$

2.3.3. Moda

Para determinarmos a moda de uma distribuição de dados, devemos observar sua distribuição de frequência absoluta.

Notas	Frequência absoluta (f_i)
0,0	1
1,0	1
2,0	2
3,0	1
4,0	1
5,0	5

6,0	4
7,0	4
8,0	5
9,0	4
10,0	2
Somatório (Σ)	30

Pela definição de moda apresentada anteriormente, temos que a distribuição utilizada para os exemplos possui duas modas que são as notas cinco e oito, pois possuem as maiores frequências absolutas na distribuição.

2.4. Medidas de Dispersão

2.4.1. Variância (σ^2 para população e S^2 para amostra)

A variância dos dados quantitativos observados em uma população é a média aritmética dos quadrados das diferenças entre cada valor observado e a média aritmética dos dados da população.

As fórmulas para o cálculo da variância populacional e da variância amostral são respectivamente:

$$\sigma^2 = \frac{\sum(X_i - \bar{X})^2}{N}$$

Onde:

σ^2 é a variância populacional;

X_i é cada valor observado na população;

\bar{X} é a média populacional;

N é o tamanho da população.

e

$$S^2 = \frac{\sum(x_i - \bar{x})^2}{n - 1}$$

S^2 é a variância amostral;

x_i é cada valor observado na amostra;

\bar{x} é a média amostral;

n é o tamanho da amostra.

O principal inconveniente na utilização da variância é o fato de fornecer medidas em quadrado das unidades originais, mas isso pode ser corrigido facilmente calculando-se sua raiz quadrada. A esse novo resultado chamamos de Desvio-padrão.

2.4.2. Desvio-padrão (σ para população e S para amostra)

O desvio-padrão é a raiz quadrada da média aritmética dos quadrados das respectivas diferenças entre cada dado observado e a média aritmética dos valores observados. Objetivamente, o desvio padrão é obtido através do cálculo da raiz quadrada da variância.

Para determinar o desvio-padrão amostral através de sua fórmula $s = \sqrt{\frac{\sum(x - \bar{x})^2}{n - 1}}$, podemos adotar o processo abaixo seguindo cada passo descrito.

Passo 1: Calcular a média \bar{x} das notas dos alunos.

Passo 2: Subtrair a média de cada valor individualmente ($x - \bar{x}$).

Passo 3: Elevar ao quadrado cada uma das diferenças obtidas no Passo 2 [$(x - \bar{x})^2$].

Passo 4: Somar todos os quadrados obtidos no Passo 3 [$\sum(x - \bar{x})^2$].

Passo 5: Dividir o valor encontrado na soma do Passo 4 por $n - 1$, isto é, uma unidade a menos que o total de notas observadas $\left[\frac{\sum(x-\bar{x})^2}{n-1}\right]$.

Passo 6: Extrair a raiz quadrada do valor obtido no Passo 5 $\left[\sqrt{\frac{\sum(x-\bar{x})^2}{n-1}}\right]$.

Com o objetivo de tornar mais claro o processo acima, efetuaremos, a título de exemplo, todos os 6 passos utilizando um conjunto de 10 notas obtidas aleatoriamente e arrumadas em ordem crescente.

Notas (x)
2,0
3,5
4,5
4,5
5,0
5,0
6,5
7,0
9,0
10,0

Passo 1: Calculando a média.

$$\bar{x} = \frac{2,0 + 3,5 + 4,5 + 4,5 + 5,0 + 5,0 + 6,5 + 7,0 + 9,0 + 10,0}{10} = 5,7$$

Passo 2: Calculando as diferenças.

Notas (x)	$(x - \bar{x})$
2,0	$2,0 - 5,7 = - 3,7$
3,5	$3,5 - 5,7 = - 2,2$
4,5	$4,5 - 5,7 = - 1,2$
4,5	$4,5 - 5,7 = - 1,2$
5,0	$5,0 - 5,7 = - 0,7$
5,0	$5,0 - 5,7 = - 0,7$
6,5	$6,5 - 5,7 = 0,8$
7,0	$7,0 - 5,7 = 1,3$
9,0	$9,0 - 5,7 = 3,3$
10,0	$10,0 - 5,7 = 4,3$

Passo 3: Calculando o quadrado de cada diferença.

Notas (x)	$(x - \bar{x})$	$(x - \bar{x})^2$
2,0	$2,0 - 5,7 = - 3,7$	$(- 3,7)^2 = 13,69$
3,5	$3,5 - 5,7 = - 2,2$	$(- 2,2)^2 = 4,84$
4,5	$4,5 - 5,7 = - 1,2$	$(- 1,2)^2 = 1,44$
4,5	$4,5 - 5,7 = - 1,2$	$(- 1,2)^2 = 1,44$
5,0	$5,0 - 5,7 = - 0,7$	$(- 0,7)^2 = 0,49$

5,0	$5,0 - 5,7 = -0,7$	$(-0,7)^2 = 0,49$
6,5	$6,5 - 5,7 = 0,8$	$(0,8)^2 = 0,64$
7,0	$7,0 - 5,7 = 1,3$	$(1,3)^2 = 1,69$
9,0	$9,0 - 5,7 = 3,3$	$(3,3)^2 = 10,89$
10,0	$10,0 - 5,7 = 4,3$	$(4,3)^2 = 18,49$

Passo 4: Calculando a soma dos quadrados das diferenças.

$$\begin{aligned} \sum (x - \bar{x})^2 &= 13,69 + 4,84 + 1,44 + 1,44 + 0,49 + 0,49 + 0,64 + 1,69 + 10,89 + 18,49 \\ &= 54,1 \end{aligned}$$

Passo 5: Dividindo a soma dos quadrados das diferenças por $n - 1$.

$$\frac{\sum (x - \bar{x})^2}{n - 1} = \frac{54,1}{10 - 1} = \frac{54,1}{9} = 6,01$$

Passo 6: Extraindo a raiz quadrada do resultado obtido no passo 5.

$$\sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}} = \sqrt{6,01} \cong 2,45$$

Ou seja, o desvio-padrão das 10 notas é igual a 2,45.

Podemos também, ao invés de executarmos manualmente o cálculo do desvio-padrão, fazê-lo através de uma planilha eletrônica do tipo Excel. Desta forma, o processo torna-se muito mais simples e mais rápido, principalmente quando a quantidade de notas é relativamente grande.

A seguir veremos o procedimento para o cálculo da média e do desvio padrão utilizando a planilha eletrônica Excel versão 2007.

Os parâmetros que contém os valores para os quais se quer calcular a média ou o desvio-padrão podem ser inseridos basicamente de duas formas: através de ponto e vírgula (;) quando se coloca célula por célula ou através de dois pontos (:) quando se coloca um intervalo de células, conforme o exemplo abaixo.

Para calcular a média, basta selecionar a célula na qual se deseja que o valor da média apareça e digitar a fórmula com os parâmetros dos valores conforme a figura abaixo.

E4		fx =MÉDIA(A2:A11)				
	A	B	C	D	E	F
1	Notas					
2	2,0					
3	3,5					
4	4,5		Média =		5,7	
5	4,5					
6	5,0		Desvio-padrão =		2,45	
7	5,0					
8	6,5					
9	7,0					
10	9,0					
11	10,0					
12						

Para calcular o desvio-padrão amostral, selecionaremos a célula na qual desejamos que o valor do desvio-padrão apareça e digitarmos a fórmula com os parâmetros dos valores conforme a figura abaixo.

E6		fx = =DESVPAD.A(A2:A11)				
	A	B	C	D	E	F
1	Notas					
2	2,0					
3	3,5					
4	4,5		Média =		5,7	
5	4,5					
6	5,0		Desvio-padrão =		2,45	
7	5,0					
8	6,5					
9	7,0					
10	9,0					
11	10,0					
12						

2.5. Escore padronizado

O escore padronizado, também chamado de escore z, representa o número de desvios-padrão pelo qual um valor x dista da média para mais ou para menos. Obtém-se esse valor através das fórmulas:

Amostra

$$z = \frac{x - \bar{x}}{s}$$

Onde:

s é o desvio-padrão amostral;

x é cada valor observado na amostra;

\bar{x} é a média amostral;

ou

População

$$z = \frac{x - \mu}{\sigma}$$

Onde:

σ é o desvio-padrão populacional;

x é cada valor observado na população;

μ é a média populacional;

“O exemplo a seguinte ilustra como os escores z podem ser usados para comparar valores, mesmo que provenham de diferentes populações.

Com a altura de 75 in (190,5 cm), Lyndon Johnson foi o presidente mais alto do último século. Com uma altura de 85 in (215,9 cm), Shaquille O’Neal é o jogador mais alto do time de basquete do Miami Heat. Quem é *relativamente* mais alto: Lyndon Johnson entre os presidentes do século passado, ou Shaquille O’Neal entre os jogadores do time de basquete do Miami Heat? Os presidentes do século passado têm alturas com média 71,5 in e desvio padrão de 2,1 in. Os jogadores de basquete do Miami Heat têm alturas com média 80 in e desvio padrão de 3,3 in.

Solução: As alturas dos presidentes e dos jogadores de basquete são provenientes de populações bem diferentes, de modo que uma comparação requer que padronizemos as alturas, convertendo-as em escore z .

$$\text{Lyndon Johnson: } z = \frac{x - \mu}{\sigma} = \frac{75 - 71,5}{2,1} = 1,67$$

$$\text{Shaquille O’Neal: } z = \frac{x - \mu}{\sigma} = \frac{85 - 80,0}{3,3} = 1,52$$

Interpretação: A altura de Lyndon Johnson está 1,67 desvios padrões acima da média, e a altura de Shaquille O’Neal está 1,52 desvios padrões acima da

média. A altura de Lyndon Johnson entre os presidentes do século passado é relativamente mais alta do que a altura de Shaquille O’Neal entre os jogadores de basquete do Miami Heat. Shaquille O’Neal é muito mais alto que Lyndon Johnson, mas Johnson é relativamente mais alto quando comparado aos colegas”. (TRIOLA, 1999)

2.6. Correlação Linear

Segundo Moore (2011, p.76 e 77), “para compreendermos uma relação estatística entre duas variáveis, medimos ambas nos mesmos indivíduos. [...] A maneira mais eficiente de se mostrar a relação entre duas variáveis quantitativas é através de um gráfico, e o mais adequado para esse fim é o diagrama de dispersão”

“Um diagrama de dispersão mostra a relação entre duas variáveis quantitativas, medidas nos mesmos indivíduos. Os valores de uma variável aparecem no eixo horizontal e os valores da outra variável aparecem no eixo vertical. Cada indivíduo nos dados aparece como o ponto no gráfico determinado pelos valores de ambas as variáveis para tal indivíduo”. (MOORE, 2011, p.77)

De acordo com Triola (1999), o coeficiente de correlação linear r mede a intensidade da relação linear entre os valores quantitativos emparelhados x e y em uma amostra. [O coeficiente de correlação linear é, algumas vezes, chamado de coeficiente de correlação do produto de momentos de Pearson, em homenagem a Karl Pearson (1857-1936), que o desenvolveu originalmente]

Abaixo, temos a fórmula para calcular o coeficiente de correlação linear:

$$r = \frac{\sum[(x_i - \bar{x})(y_i - \bar{y})]}{\sqrt{(\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2)}}$$

Porém, recorreremos à função da planilha eletrônica Excel para efetuar os cálculos tendo em vista sua praticidade.

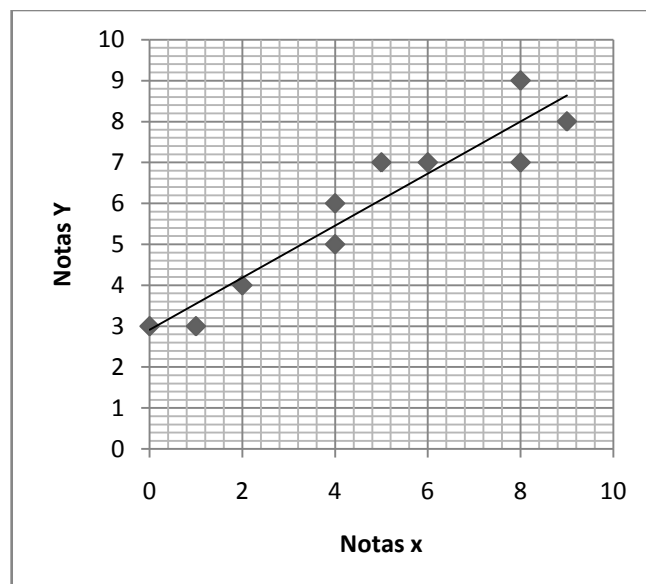
Exemplos de Cálculo do Coeficiente de correlação Linear

Exemplo 1

Considere uma turma com dez alunos numerados de um a dez. Chamaremos de notas x as notas obtidas por estes alunos em uma disciplina qualquer e notas y as notas obtidas por estes mesmos alunos em outra disciplina diferente da primeira.

Nº do Aluno	Notas x	Notas y
1	4	5
2	5	7
3	8	9
4	1	3
5	4	6
6	2	4
7	6	7
8	8	7
9	0	3
10	9	8

Antes de calcularmos o coeficiente de correlação linear, construiremos o gráfico de dispersão das notas para que façamos uma análise inicial.



De acordo com o gráfico acima, podemos observar que à medida que a nota na primeira disciplina (nota x) aumenta a nota na segunda disciplina (nota y) também aumenta.

A figura abaixo mostra a utilização da função CORREL da planilha eletrônica Excel para realizar o cálculo do coeficiente de correlação linear.

F4		fx		=CORREL(C4:C13;D4:D13)					
	A	B	C	D	E	F	G	H	
1									
2									
3		Nº do Aluno	Notas x	Notas Y		Coeficiente de Correlação Linear			
4		1	4	5		0,945172223			
5		2	5	7					
6		3	8	9					
7		4	1	3					
8		5	4	6					
9		6	2	4					
10		7	6	7					
11		8	8	7					
12		9	0	3					
13		10	9	8					
14									

Exemplo 2

Utilizaremos a mesma ideia utilizada no exemplo 1, porém modificando os valores das notas a fim de verificar a diferença entre os respectivos gráficos de dispersão e coeficientes de variação.

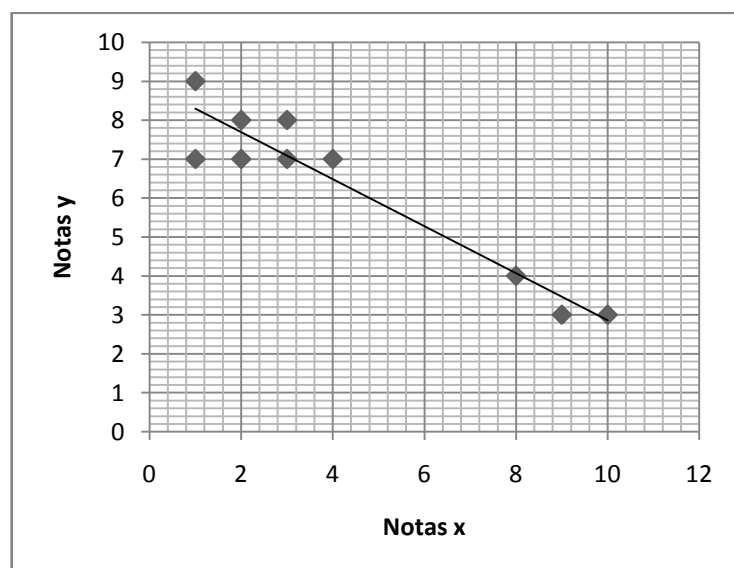
Nº do Aluno	Notas x	Notas Y
1	10	5
2	5	7
3	8	9
4	5	6
5	9	8
6	2	7
7	6	7

Exemplo 3

Utilizaremos a mesma ideia utilizada nos exemplos 1 e 2, porém modificando os valores das notas a fim de verificar a diferença entre os respectivos gráficos de dispersão e coeficientes de variação.

Nº do Aluno	Notas x	Notas Y
1	9	3
2	1	7
3	8	4
4	10	3
5	4	7
6	2	8
7	3	7
8	1	9
9	2	7
10	3	8

O gráfico de dispersão:



O cálculo do coeficiente de correlação:

F4		fx =CORREL(C4:C13;D4:D13)						
	A	B	C	D	E	F	G	H
1								
2								
3		Nº do Aluno	Notas x	Notas Y		Coeficiente de Correlação Linear		
4		1	9	3		-0,950132629		
5		2	1	7				
6		3	8	4				
7		4	10	3				
8		5	4	7				
9		6	2	8				
10		7	3	7				
11		8	1	9				
12		9	2	7				
13		10	3	8				
14								

Observações:

- Os gráficos obtidos nos exemplos 1, 2 e 3, nos mostram “tendências” de que exista respectivamente, correlação positiva, nula (ou não existe correlação) e negativa;

- Para auxiliar na percepção das “tendências” citadas anteriormente, utilizamos o layout do gráfico de dispersão (LAYOUT 3) que contém uma reta de tendência.

- A reta descrita na observação anterior é chamada de reta de regressão (ou reta de melhor ajuste ou reta de mínimos quadrados). A reta de regressão é a que melhor se ajusta aos dados amostrais e o critério específico usado para se determinar qual reta se ajusta melhor é a propriedade dos mínimos quadrados;

- Comentários pertinentes a valores do coeficiente de correlação linear serão realizados no capítulo 5;

- Os diagramas de dispersão foram gerados através das ferramentas de gráficos do Excel seguindo os passos abaixo:

Passo 1: Selecione os dados que deverão compor o diagrama de dispersão. É conveniente que esses dados estejam em colunas adjacentes.

Passo 2: Clique em **INSERIR » GRÁFICO » DISPERSÃO**

Passo 3: Para que o gráfico apresente a reta que mais se aproxima da tendência dos pontos do diagrama de dispersão, dê um duplo clique com o botão esquerdo do mouse e selecione em **LAYOUT DO GRÁFICO** o **LAYOUT 3**.

The screenshot shows the Microsoft Excel interface with the 'Inserir' (Insert) tab selected. The 'Gráficos' (Charts) group is active, and the 'Dispersão' (Scatter) dropdown menu is open, showing various chart styles. The data table is as follows:

	A	B	C	D	E	F	G
1							
2							
3			Nº do Aluno	Notas x	Notas Y		
4			1	9	3		
5			2	1	7		
6			3	8	4		
7			4	10	3		
8			5	4	7		
9			6	2	8		
10			7	3	7		
11			8	1	9		
12			9	2	7		
13			10	3	8		
14							

Capítulo 3

A utilização do desvio-padrão

Vamos analisar a seguinte situação hipotética: duas turmas fizeram a mesma prova da mesma disciplina e obtiveram a mesma média. Será que podemos afirmar que o aprendizado foi igual nas duas turmas?

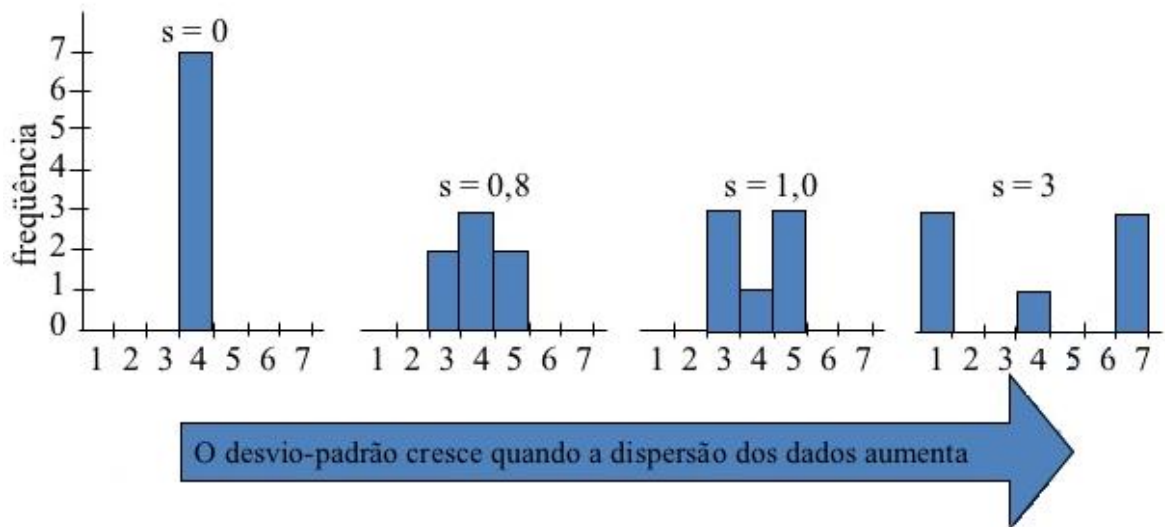
Definitivamente a resposta é não! Conhecendo apenas a média não podemos fazer tal afirmação.

Utilizar apenas valor da média para esse tipo de análise não é uma forma eficiente para caracterizar uma distribuição. Devemos, para complementar esse estudo, levar em conta um outro tipo de medida que nos permita conhecer a forma como os dados observados se distribuem, dispersam ou variam na distribuição observada.

Alguns indicadores são apropriados para medir a dispersão ou variabilidade de um conjunto de dados, como por exemplo: a amplitude total, a variância, o desvio padrão e o coeficiente de variação. O desvio-padrão é, sem dúvida, a mais importante das medidas de dispersão.

Devemos ter em mente que o desvio-padrão mede a variação entre os valores. Em uma turma em que as notas estão próximas umas das outras, encontraremos valor relativamente pequeno para o desvio-padrão. Enquanto que se as notas estiverem afastadas entre si, o valor do desvio-padrão tenderá a aumentar.

Na figura abaixo, TRIOLA (1999) mostra que apesar de possuírem a mesma média, as quatro distribuições representadas são bastante diferentes, por possuírem desvios-padrão diferentes entre si.



Abaixo, com a finalidade de ilustrar e enriquecer o exemplo acima, temos os dados em tabelas de distribuição de frequência com os respectivos valores da média e desvio-padrão. Para isso, consideramos, da esquerda para a direita, respectivamente, distribuições 1, 2, 3 e 4.

Distribuição 1	
Valor	Frequência
4	7

Distribuição 2	
Valor	Frequência
3	2
4	3
5	2

Média	4
Desv. Padrão	0

Média	4
Desv. Padrão	0,8

Distribuição 3	
Valor	Frequência
3	3
4	1
5	3

Distribuição 4	
Valor	Frequência
1	3
4	1
7	3

Média	4
Desv. Padrão	1

Média	4
Desv. Padrão	3

Mas, para que o desvio-padrão exerça seu importante papel na interpretação dos dados, é fundamental que o interessado consiga relacionar o valor obtido do desvio-padrão com os dados da série em questão.

Em nosso estudo temos por objetivo mostrar que analisar o resultado de uma avaliação somente através da média das notas obtidas, como normalmente é feito pela maioria dos educadores, não é eficiente o bastante.

Calcular a média é necessário, mas não é suficiente!

Retornaremos, agora, a situação hipotética do início deste capítulo. As turmas A e B obtiveram médias iguais em uma determinada avaliação. Vamos analisar, com o auxílio do desvio-padrão, se o aprendizado aconteceu de forma igualitária nas duas turmas em questão. Além disso, de posse da informação de que grande parte das instituições de ensino tem como critério para progressão para a série seguinte, o aproveitamento mínimo de cinquenta por cento em cada disciplina, faremos uma análise complementar nesse aspecto.

TURMA A			
Aluno	Nota	Aluno	Nota
1	5,5	16	6,0
2	4,5	17	5,5
3	5,0	18	5,0
4	7,0	19	5,5
5	6,5	20	6,0
6	7,0	21	7,0
7	5,0	22	6,0
8	3,5	23	5,0
9	6,5	24	5,5
10	6,0	25	4,5
11	6,0	26	8,5
12	5,0	27	6,0
13	5,5	28	5,5
14	6,5	29	5,5
15	7,5	30	6,0

TURMA A
MÉDIA 5,82
DESV. PADRÃO 1,00

TURMA B			
Aluno	Nota	Aluno	Nota
1	8,0	16	10,0
2	8,5	17	1,0
3	2,5	18	2,5
4	2,0	19	2,0
5	10,0	20	7,0
6	1,0	21	9,5
7	0,0	22	0,5
8	3,0	23	8,5
9	5,0	24	7,0
10	3,5	25	9,0
11	9,0	26	10,0
12	1,0	27	8,0
13	0,5	28	8,5
14	9,5	29	9,5
15	8,0	30	10,0

TURMA B
MÉDIA 5,82
DESV. PADRÃO 3,69

Apesar de ambas as turmas, A e B, terem obtido média igual a 5,82, podemos observar que os valores obtidos com o cálculo do desvio-padrão são bastante diferentes. Isso nos revela que na turma A, em que a média das notas obtidas pelos alunos foi 5,82 e o desvio-padrão foi 1,0, o aprendizado ocorreu de forma mais homogênea entre os alunos, quando comparado com a turma B, que obteve a mesma média 5,82, porém com o desvio-padrão significativamente mais alto.

Com o intuito de complementar uma análise desses resultados, observaremos o quantitativo de alunos que obtiveram rendimento maior ou igual a cinquenta por cento.

TURMA A	
Frequência de notas $\geq 5,0$	27
Frequência de notas $< 5,0$	3

TURMA B	
Frequência de notas $\geq 5,0$	18
Frequência de notas $< 5,0$	12

Analisando as informações acima, observamos que apesar das médias das duas turmas serem iguais, na turma A noventa por cento dos alunos obtiveram aproveitamento de cinquenta por cento ou mais, enquanto que na turma B, apenas sessenta por cento obtiveram aproveitamento semelhante.

Outro tipo de análise possível é estimar o percentual de alunos que obtiveram notas dentro de um intervalo desejado. Para isso, faremos uso da Regra Empírica (Regra 68-95-99) ou do Teorema de Tchebichev, dependendo da forma da distribuição.

3.1. Regra Empírica ou Regra 68-95-99

- Cerca de 68% dos valores estão a menos de 1 desvio-padrão a contar da média;
- Cerca de 95% dos valores estão a menos de 2 desvios-padrão a contar da média;
- Cerca de 99,7% dos valores estão a menos de 3 desvios-padrão a contar da média.

3.2. Teorema de Tchebichev

A proporção (ou fração) de qualquer conjunto de dados a menos de k desvios-padrão a contar da média é sempre ao menos $1 - 1/k^2$, onde k é um número positivo maior que 1. Para $k = 2$ e $k = 3$, temos os seguintes resultados específicos:

- Ao menos $3/4$ (ou 75%) de todos os valores estão no intervalo que vai de 2 desvios-padrão abaixo da média a 2 desvios-padrão acima da média ($\bar{x} - 2s$ a $\bar{x} + 2s$).
- Ao menos $8/9$ (ou 89%) de todos os valores estão no intervalo que vai de 3 desvios-padrão abaixo da média a 3 desvios-padrão acima da média ($\bar{x} - 3s$ a $\bar{x} + 3s$).

De acordo com Triola (1999), a regra empírica (68-95-99) se aplica apenas a conjunto de dados com distribuição em forma de sino (uma distribuição normal ou outra distribuição que possa ser aproximada pela a distribuição normal). O Teorema de Tchebichev se aplica a qualquer conjunto de dados, mas seus resultados são muito aproximados.

Em geral, essas ferramentas são bastante utilizadas quando se tem um quantitativo elevado de dados, em nosso contexto, uma quantidade muito grande de notas. Dessa forma, pode-se estimar o percentual de notas que compõem um determinado intervalo.

Apesar do pequeno número de alunos, tomaremos como exemplo a turma A e analisaremos sob os parâmetros do Teorema de Tchebichev. Isto feito, podemos apontar as seguintes características:

- Fazendo $k = 1,18$; ao menos 28,2% dos alunos obtiveram nota no intervalo (4,64; 7,0);
- Fazendo $k = 2$, ao menos 75% dos alunos obtiveram nota no intervalo (3,82; 7,82);
- Fazendo $k = 3$, ao menos 89% dos alunos obtiveram nota no intervalo (2,82; 8,82).

Capítulo 4

Escore padronizado, como utilizar?

Ao final da correção de uma prova, ela recebe uma nota bruta. A padronização consiste em mudar a escala tendo como referências a média e o desvio padrão de cada prova. A padronização evita, por exemplo, que uma prova muito difícil apresente resultados incoerentes quando comparados aos resultados de outra prova mais fácil.

Segundo Triola (1999), “a importância dos escores z na estatística reside no fato de que eles permitem distinguir entre valores usuais e valores raros, ou incomuns.

“Exemplo: Uma professora de estatística aplica dois testes diferentes a duas turmas do seu curso. Os resultados foram

Turma 1: $\bar{x} = 75$ e $s = 14$

Turma 2: $\bar{x} = 40$ e $s = 8$

Que nota é relativamente melhor: 82 no teste da Turma 1, ou 46 no da turma 2?

Solução: Não podemos comparar diretamente as notas 82 e 46 porque provêm de escalas diferentes. Transformamo-las, portanto, em escores z . Para o valor 82 da Turma 1, obtemos o escore z 0,50, porque

$$z = \frac{x - \bar{x}}{s} = \frac{82 - 75}{14} = 0,50$$

Para a nota 46 da Turma 2, o escore z correspondente é 0,75, porque

$$z = \frac{x - \bar{x}}{s} = \frac{46 - 40}{8} = 0,75$$

Isso significa que a nota 82 do teste da Turma 1 está 0,5 desvio-padrão acima da média, enquanto a nota 46 do teste da turma 2 está 0,75 desvio-padrão acima da média. Isso implica que o resultado 46 do teste da Turma 2 é melhor, relativamente. Embora inferior a nota 82, a nota 46 tem melhor posição relativa no contexto dos outros resultados do teste”. (Triola, 1999, p.46)

Com o objetivo de exemplificar a utilização do escore padronizado, recorreremos mais uma vez a uma situação hipotética.

Considere que as turmas A, B e C, foram submetidas a provas de matemática de diferentes professores. Suponha agora, que tenhamos que escolher 6 dentre os 30 alunos para representar a escola em uma Olimpíada de Matemática. E agora, qual o melhor critério a se utilizar?

Turma A		Turma B		Turma C	
Alunos	Notas	Alunos	Notas	Alunos	Notas
1A	7	1B	10	1C	10
2A	7	2B	9	2C	10
3A	6	3B	7	3C	9
4A	4	4B	7	4C	8
5A	3	5B	7	5C	5
6A	2	6B	7	6C	4
7A	2	7B	5	7C	4
8A	1	8B	5	8C	3
9A	1	9B	3	9C	2
10A	0	10B	0	10C	1

Média	3,3	Média	6	Média	5,6
-------	-----	-------	---	-------	-----

Desv. P	2,58	Desv. P	2,91	Desv. P	3,37
---------	------	---------	------	---------	------

Certamente, algumas pessoas escolheriam os 6 alunos que obtiveram as maiores notas brutas. Outras escolheriam os dois melhores alunos de cada turma. Será que estas são as melhores opções?

Temos que levar em conta que os alunos foram avaliados por diferentes provas e por diferentes professores. Para minimizar a interferência desses fatores na escolha, o mais adequado a fazer é não realizar a escolha observando as notas brutas, mas sim após realizar a padronização das notas.

Conforme descrito no capítulo 2, utilizaremos a fórmula $z = \frac{x - \bar{x}}{s}$, para o cálculo da nota padronizada para cada aluno. Este procedimento poderá ser feito manualmente, mas a utilização de uma planilha eletrônica o torna mais rápido e preciso.

Turma A			Turma B			Turma C		
Alunos	Notas	z	Alunos	Notas	z	Alunos	Notas	z
1A	7	1,432	1B	10	1,376	1C	10	1,304
2A	7	1,432	2B	9	1,032	2C	10	1,304
3A	6	1,045	3B	7	0,344	3C	9	1,008
4A	4	0,271	4B	7	0,344	4C	8	0,712
5A	3	-0,116	5B	7	0,344	5C	5	-0,178
6A	2	-0,503	6B	7	0,344	6C	4	-0,474
7A	2	-0,503	7B	5	-0,344	7C	4	-0,474
8A	1	-0,890	8B	5	-0,344	8C	3	-0,771
9A	1	-0,890	9B	3	-1,032	9C	2	-1,067
10A	0	-1,277	10B	0	-2,065	10C	1	-1,364

Observando as notas das três turmas, constatamos que a maior nota padronizada, não é necessariamente, a maior nota bruta. Os alunos 1A e 2A, por exemplo, possuem nota bruta inferior às notas dos alunos 1B, 2B, 1C, 2C, 3C, e 4C, mas quando observadas a luz do escore padrão, as notas dos alunos 1A e 2A, são relativamente maiores que as dos outros alunos.

Note que a escolha dos alunos se difere à medida que estabelecemos um novo critério.

Critérios de Escolha	Alunos Escolhidos
2 maiores notas brutas de cada turma	1A, 2A, 1B, 2B, 1C, 2C
6 maiores notas brutas	1B, 2B, 1C, 2C, 3C, 4C
6 maiores escores padronizados	1A, 2A, 3A, 1B, 1C, 2C

Primeiramente, partiremos da hipótese de que os alunos estão distribuídos de forma aleatória nas turmas, ou seja, os alunos não foram agrupados segundo qualquer critério que quantifique sua capacidade.

Sabendo que as provas foram elaboradas e corrigidas por diferentes professores, poderíamos inferir, por exemplo, que a prova a que a turma A foi submetida, possui maior grau de dificuldade que as outras. E, portanto, comparar as notas brutas não seria um critério

justo. Esse é um dos motivos pelos quais se acredita que a utilização do escore padronizado seja uma forma “relativamente neutra” para decidir sobre situações como essa.

Com o objetivo de tornar os escores padronizados com uma melhor forma de representação e conseqüentemente, maior facilidade na interpretação, podemos recorrer aos recursos normalmente utilizados por comissões organizadoras de vestibulares e concursos públicos.

Em um documento, que explica como é feita a padronização das notas do seu vestibular, a Fundação Getúlio Vargas nos apresenta o seguinte:

Vamos ilustrar isso com um exemplo:

A média das notas em História foi 6,25, em uma sala com 6 alunos, e as notas desses alunos foram as seguintes:

4,8 – 5,5 – 6,5 – 6,5 – 7,0 – 7,2
(repare que a média dessas 6 notas é igual a 6,25)

A média das notas em Matemática foi 5,0, em uma sala com 6 alunos, e as notas desses alunos foram as seguintes:

2,0 – 3,0 – 4,0 – 6,0 – 7,0 – 8,0
(repare que a média dessas 6 notas é igual a 5,0).

[...]

A padronização considera a média das notas e o desvio-padrão, buscando converter todas as notas para uma mesma escala, com o objetivo de tornar, como função de seu desempenho global e não absoluto, a classificação do candidato.

A padronização da nota de uma disciplina, no vestibular da FGV, é obtida por meio da seguinte fórmula:

$NP_i = 5 + [(N_i - M_x) / DP_x]$, na qual;

NP_i = nota padronizada do candidato *i* naquela prova;

N_i = nota bruta do candidato *i* (antes da padronização) naquela prova;

M_x = média das notas dos candidatos naquela prova,

DP_x = desvio-padrão das notas naquela prova.

Onde:

NP = Nota Padronizada

N = Nota bruta do candidato

M = Média geral das notas brutas da prova

DP = Desvio Padrão das notas brutas da prova

Tomando-se os números mencionados acima, as notas padronizadas seriam:

História:

Nota bruta do candidato = 7,0

Média geral das notas brutas = 6,25

Desvio-padrão das notas brutas da prova = 0,84

Logo, a nota padronizada do candidato será

$$NP = 5 + (7 - 6,25)/0,84 = 5,89$$

Matemática:

Nota bruta do candidato = 7,0

Média geral das notas brutas = 5,0

Desvio-padrão das notas brutas da prova = 2,16

Logo, a nota padronizada do candidato será

$$NP = 5 + (7 - 5)/2,16 = 5,93$$

(FGV, Nota padronizada)

Dessa forma, utilizando o mesmo modelo de padronização utilizado pela FGV (somando 5 unidades a cada valor de z encontrado) nosso exemplo se apresenta da seguinte forma:

Turma A			
Alunos	Notas	z	FGV
1ª	7	1,432	6,432
2ª	7	1,432	6,432
3ª	6	1,045	6,045
4ª	4	0,271	5,271
5ª	3	-0,116	4,884
6ª	2	-0,503	4,497
7ª	2	-0,503	4,497
8ª	1	-0,890	4,110
9ª	1	-0,890	4,110
10ª	0	-1,277	3,723

Turma B			
Alunos	Notas	z	FGV
1B	10	1,376	6,376
2B	9	1,032	6,032
3B	7	0,344	5,344
4B	7	0,344	5,344
5B	7	0,344	5,344
6B	7	0,344	5,344
7B	5	-0,344	4,656
8B	5	-0,344	4,656
9B	3	-1,032	3,968
10B	0	-2,065	2,935

Turma C			
Alunos	Notas	z	FGV
1C	10	1,304	6,304
2C	10	1,304	6,304
3C	9	1,008	6,008
4C	8	0,712	5,712
5C	5	-0,178	4,822
6C	4	-0,474	4,526
7C	4	-0,474	4,526
8C	3	-0,771	4,229
9C	2	-1,067	3,933
10C	1	-1,364	3,636

Observações:

- A coluna denominada de FGV apresenta os valores das respectivas notas segundo o modelo utilizado pela instituição;

- Para o caso de termos notas que variem entre 0 (zero) e 100 (cem), inclusive, a padronização mais coerente a se fazer é multiplicar o valor do escore z encontrado por 10 e somar o resultado a 50.

Capítulo 5

Existe relação entre as notas das disciplinas?

Será que os alunos que tiram boas notas em Matemática também tiram boas notas em Física? E se compararmos as notas nas disciplinas História e Geografia, será que isso acontece?

A motivação para a realização do estudo contemplado nesse capítulo nasceu da idéia de que algumas disciplinas têm muitas “coisas em comum”, portanto, seria razoável supor que os alunos que tiram notas altas em uma delas, também devem tirar notas altas na outra. Para o nosso critério de análise utilizamos como disciplinas que possuem as tais “coisas em comum” as seguintes: Português (Port) e Literatura (Lit); Matemática (Mat) e Física (Fís); História (His) e Geografia (Geo); Biologia (Bio) e Química (Quí).

Para respondermos a estas e outras perguntas que frequentemente nos deparamos recorreremos a um conceito estatístico, a correlação linear. E, através do coeficiente de correlação de Pearson, quantificaremos as relações caso existam.

Como base de dados para nossa pesquisa, utilizaremos as notas obtidas no ano de 2014, pelos 43 alunos de uma turma do 2º ano do ensino médio de um colégio de classe média, localizado no bairro Freguesia em Jacarepaguá – RJ, a fim de chegarmos às respostas das perguntas do início deste capítulo. Essas notas variam em um intervalo de 0 (zero) a 100 (cem), inclusive. Por motivos éticos, os nomes dos alunos não serão divulgados no neste trabalho. Sendo assim, cada aluno será representado apenas por um número.

Com o objetivo de tornar o estudo mais minucioso, calcularemos o coeficiente de correlação entre as notas das disciplinas em 5 épocas distintas do ano letivo: 1º bimestre, 2º bimestre, 3º bimestre, 4º bimestre e média final. Começaremos pelas notas do 1º bimestre, realizando uma análise detalhada e explicando o método utilizado. Em seguida apresentaremos, respectivamente, as notas e as correlações dos bimestres subsequentes.

Notas do 1º Bimestre (Turma 2011)								
Aluno	Port	Lit	Mat	Fís	His	Geo	Bio	Quí
1	60	95	20	65	50	65	40	50
2	50	80	30	60	50	60	65	50
3	65	100	60	50	50	70	45	65
4	50	20	35	70	55	50	50	60
5	50	80	50	80	70	60	30	70
6	50	90	30	70	70	35	50	73
7	50	100	60	60	50	70	35	63
8	50	60	55	80	50	45	25	50
9	60	100	55	70	50	80	25	65
10	50	95	30	60	60	45	20	65
11	50	100	20	60	55	50	30	50
12	50	90	75	75	55	70	35	50
13	55	30	0	70	60	75	60	50
14	65	70	25	60	50	45	55	63
15	50	80	30	80	70	70	40	50
16	50	25	20	80	50	60	20	50
17	60	85	75	70	75	65	55	50
18	50	75	30	60	70	40	45	65
19	50	60	40	60	60	45	55	50
20	75	100	70	60	70	60	55	58
21	50	85	60	75	70	50	30	58
22	85	100	70	50	90	90	70	80
23	50	75	25	75	80	50	35	80
24	50	85	20	65	65	60	25	50
25	50	90	20	65	70	55	25	70
26	45	85	40	70	60	60	30	70
27	35	100	50	70	70	65	25	58
28	50	100	25	60	70	50	35	50
29	50	85	25	60	70	40	35	50
30	60	95	20	60	50	75	20	50
31	50	80	25	60	40	60	50	55
32	50	100	30	60	60	40	25	50
33	50	75	40	60	80	40	30	50
34	50	80	35	55	60	40	30	55
35	50	90	60	60	70	80	50	53
36	50	70	35	50	60	65	50	50
37	50	75	35	50	60	25	25	53
38	60	85	35	60	55	50	55	55
39	55	85	20	50	60	75	10	50
40	60	100	35	60	60	60	45	55
41	55	95	30	60	50	30	30	55
42	40	35	50	25	52	55	55	80

43	60	60	80	40	30	65	70	20
----	----	----	----	----	----	----	----	----

Inicialmente, iremos construir os gráficos de dispersão utilizando as notas, de acordo com as disciplinas consideradas de mesma área de conhecimento. Os pontos no gráfico representam pares de notas de cada um dos alunos da turma, onde as coordenadas de cada ponto são as notas obtidas pelo aluno nas disciplinas em questão.

Devemos estudar o padrão global dos pontos. E, se houver um padrão, devemos observar a sua direção.

Diagrama de dispersão das notas obtidas em Português e Literatura.

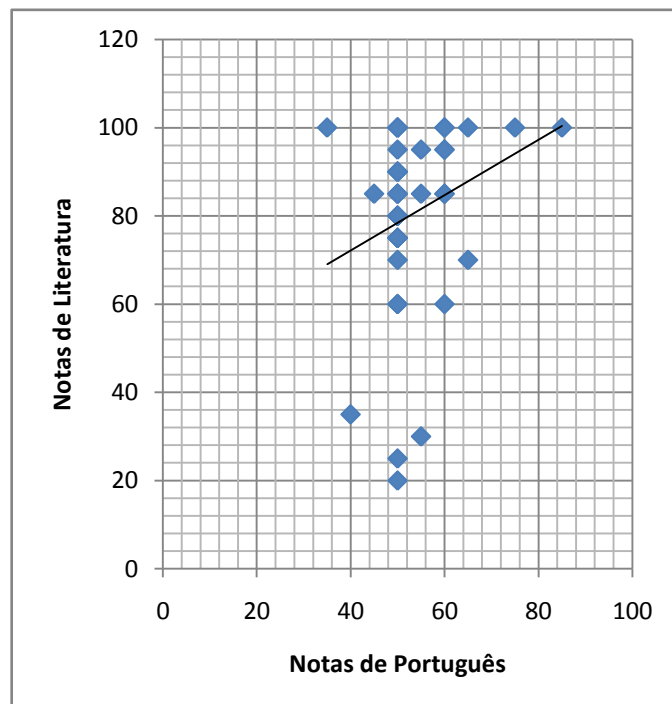


Diagrama de dispersão das notas obtidas em Matemática e Física.

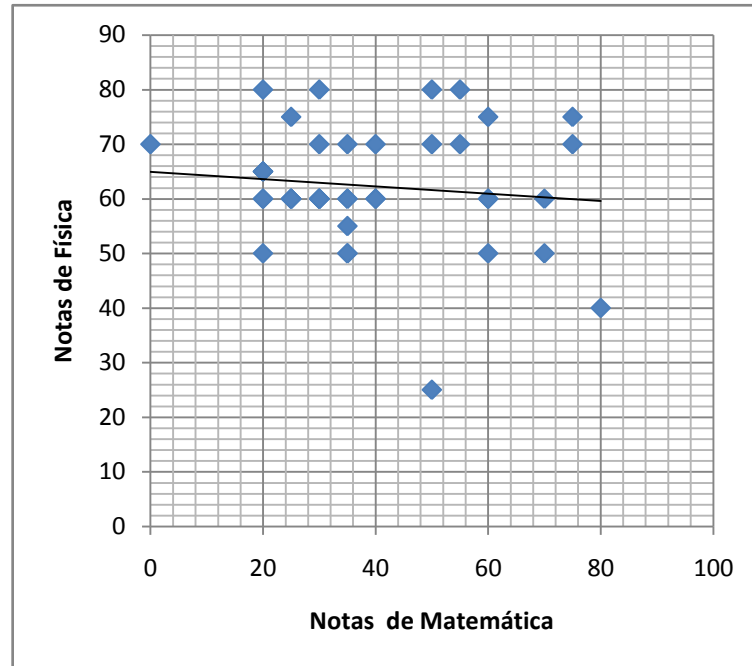


Diagrama de dispersão das notas obtidas em História e Geografia.

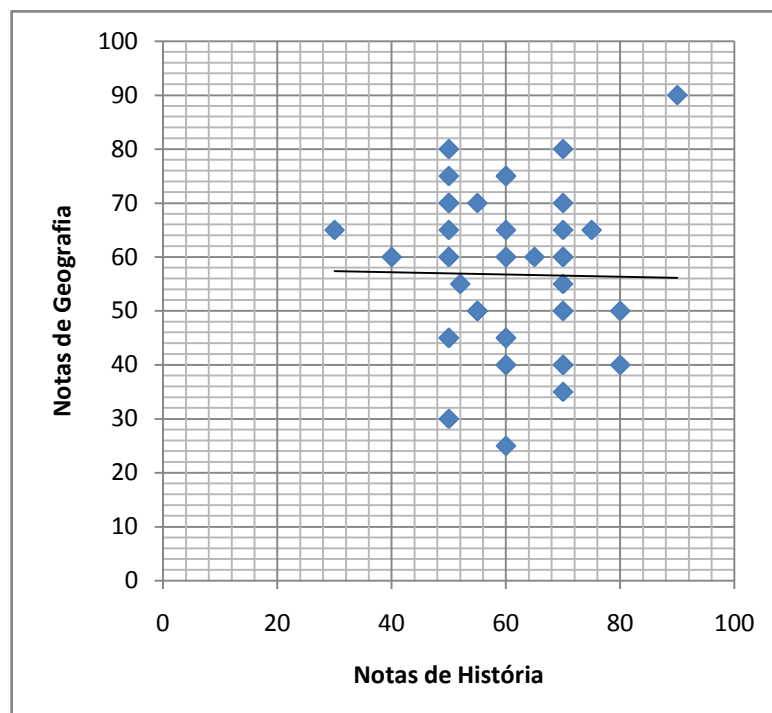
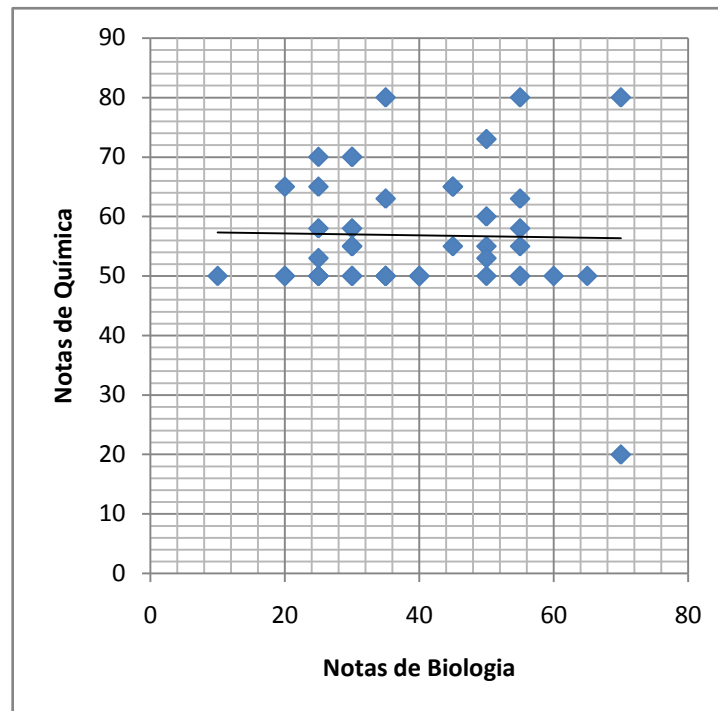


Diagrama de dispersão das notas obtidas em Biologia e Química.



Como sabemos que o exame visual é bastante subjetivo, lançaremos mão da utilização do coeficiente de correlação linear de Pearson para que possamos alcançar conclusões mais satisfatórias.

Abaixo, temos os respectivos resultados dos cálculos realizados para o coeficiente de correlação linear entre os pares de disciplinas denominadas de mesma área de conhecimento.

Correlações entre Disciplinas da mesma área			
Port x Lit	0,252	His x Geo	-0,016
Mat x Fís	-0,114	Bio x Quí	-0,023

De acordo com os parâmetros encontrados em grande parte da literatura pesquisada, os valores observados acima não são suficientes para afirmarmos que exista correlação linear entre as disciplinas testadas.

Porém, no intuito de tornar a pesquisa mais precisa, utilizaremos a tabela de “Valores Críticos do Coeficiente de Correlação de Pearson (r)” adotada por Triola (2008, p.626).

A utilização da tabela consiste em verificar, para o nível de significância (α) desejado, se o valor do módulo do coeficiente de correlação calculado excede o valor encontrado na tabela para o respectivo tamanho da amostra. Caso o valor exceda, concluímos que há uma correlação linear. Caso contrário, não há evidência suficiente para apoiar uma conclusão de uma correlação linear.

n	$\alpha=0,05$	$\alpha=0,01$
4	0,950	0,999
5	0,878	0,959
6	0,811	0,917
7	0,754	0,875
8	0,707	0,834
9	0,666	0,798
10	0,632	0,765
11	0,602	0,735
12	0,576	0,708
13	0,553	0,684
14	0,532	0,661
15	0,514	0,641
16	0,497	0,623
17	0,482	0,606
18	0,468	0,590
19	0,456	0,575
20	0,444	0,561
25	0,396	0,505
30	0,361	0,463
35	0,335	0,430
40	0,312	0,402
45	0,294	0,378
50	0,279	0,361
60	0,254	0,330
70	0,236	0,305
80	0,220	0,286
90	0,207	0,269
100	0,196	0,256

A tabela não possui os valores críticos para o tamanho da amostra que utilizamos em nossa pesquisa, por isso, através de interpolação linear, estimou-se os valores abaixo.

n	$\alpha=0,05$	$\alpha=0,01$
40	0,312	0,402
41	0,308	0,397
42	0,305	0,392
43	0,301	0,388
44	0,298	0,383
45	0,294	0,378

Interpretando os valores críticos encontrados temos que, partindo da hipótese de que não exista correlação linear entre as disciplinas, para uma turma com 43 alunos (amostra de tamanho 43), temos 0,05 de probabilidade do módulo do valor do coeficiente de correlação linear calculado entre duas disciplinas exceder 0,301 e 0,01 de probabilidade dele exceder 0,388.

Adotaremos em nosso trabalho o nível de significância de 5% para analisarmos os resultados obtidos. Para tornar a visualização mais fácil, utilizaremos a fonte na cor vermelha para valores do coeficiente de correlação maiores ou iguais do que 0,301.

Abaixo temos os valores calculados para o coeficiente de correlação entres as notas obtidas pelos alunos nas outras disciplinas.

Outras Correlações (1º Bimestre)			
Port x Mat	0,280	Mat x His	0,044
Port x Fís	-0,167	Mat x Geo	0,295
Port x His	0,076	Mat x Bio	0,306
Port x Geo	0,360	Mat x Quí	-0,001
Port x Bio	0,413	Fís x His	0,235
Port x Quí	0,035	Fís x Geo	0,015
Lit x Mat	0,188	Fís x Bio	-0,326
Lit x Fís	-0,009	Fís x Quí	0,038
Lit x His	0,233	His x Bio	-0,043
Lit x Geo	0,112	His x Quí	0,434
Lit x Bio	-0,235	Geo x Bio	0,201
Lit x Quí	0,096	Geo x Quí	-0,006

Podemos observar que os resultados do primeiro bimestre apontam uma fraca correlação entre alguns pares de disciplinas.

Notas e correlações do 2º Bimestre

Notas do 2º Bimestre (Turma 2011)								
Aluno	Port	Lit	Mat	Fís	Hist	Geo	Bio	Quí
1	70	80	45	75	60	60	55	50
2	80	95	50	60	60	65	55	60
3	80	85	65	70	75	75	75	100
4	65	85	65	50	60	65	50	70
5	70	80	60	80	50	80	60	75
6	70	50	55	50	50	50	35	40
7	75	80	35	70	60	55	55	70
8	60	70	20	90	60	65	65	55
9	75	80	80	80	60	90	50	65
10	55	80	50	70	60	60	45	45
11	60	80	35	75	60	65	45	60
12	55	0	75	75	60	70	50	60
13	60	60	20	70	50	75	60	55
14	70	55	45	90	60	80	55	60
15	80	50	50	90	60	60	35	60
16	65	50	20	90	30	60	35	15
17	65	100	65	70	60	75	70	80
18	65	80	50	70	60	65	45	70
19	60	55	20	70	60	50	45	60
20	65	80	60	70	60	65	55	60
21	65	80	40	90	60	70	65	0
22	80	70	85	65	60	80	80	95
23	60	75	45	90	70	65	35	45
24	80	50	45	60	60	55	35	55
25	80	85	60	70	30	50	25	25
26	70	90	50	50	60	65	50	40
27	70	85	75	75	40	100	60	70
28	70	80	50	80	50	75	50	60
29	65	100	45	60	60	65	45	70
30	80	65	45	60	60	75	35	55
31	85	90	25	60	60	60	65	50
32	60	50	45	60	0	60	0	60
33	65	90	45	60	40	65	15	55
34	80	80	50	65	60	50	35	40
35	65	80	50	55	50	45	70	60
36	65	50	35	40	0	70	45	40
37	60	85	40	40	60	55	45	60
38	70	90	45	60	40	70	60	60

39	70	55	45	40	50	70	40	60
40	65	90	45	50	50	70	60	60
41	60	75	45	60	50	75	45	55
42	50	56	60	85	72	65	75	75
43	60	60	80	50	60	68	84	64

Correlações entre Disciplinas da mesma área			
Port x Lit	0,250	His x Geo	0,019
Mat x Fís	-0,060	Bio x Quí	0,393

Outras Correlações (2º Bimestre)			
Port x Mat	0,125	Mat x His	0,184
Port x Fís	-0,070	Mat x Geo	0,416
Port x His	0,070	Mat x Bio	0,280
Port x Geo	0,024	Mat x Quí	0,452
Port x Bio	-0,029	Fís x His	0,269
Port x Quí	0,059	Fís x Geo	0,219
Lit x Mat	0,012	Fís x Bio	0,095
Lit x Fís	-0,125	Fís x Quí	-0,109
Lit x His	0,186	His x Bio	0,490
Lit x Geo	0,080	His x Quí	0,313
Lit x Bio	0,195	Geo x Bio	0,335
Lit x Quí	0,126	Geo x Quí	0,361

Notas e correlações do 3º Bimestre

Notas do 3º Bimestre (Turma 2011)								
Aluno	Port	Lit	Mat	Fís	His	Geo	Bio	Quí
1	70	60	40	60	65	40	60	65
2	75	80	35	60	50	50	40	55
3	65	90	70	60	80	70	80	70
4	70	80	60	70	80	95	65	55
5	55	40	45	65	75	70	55	75
6	45	60	50	50	60	70	40	45

7	60	70	55	70	65	75	70	45
8	55	70	45	60	70	60	50	40
9	75	85	65	70	75	75	90	100
10	80	65	35	70	40	35	30	60
11	35	40	25	60	75	40	30	55
12	40	60	55	60	50	60	50	65
13	50	60	30	60	40	50	60	65
14	45	80	90	90	60	85	45	55
15	80	60	25	70	70	40	55	60
16	55	50	25	60	50	45	50	30
17	75	80	70	60	50	60	70	60
18	55	90	45	70	60	75	60	65
19	45	85	55	50	50	60	80	80
20	50	75	90	60	90	80	85	65
21	55	70	50	60	50	60	45	55
22	75	100	80	65	75	95	90	95
23	65	75	30	60	30	30	45	30
24	60	90	30	60	30	40	50	65
25	50	50	40	60	80	60	45	60
26	0	0	0	70	0	0	0	0
27	0	0	0	70	0	0	0	0
28	70	80	50	60	65	60	55	70
29	40	70	65	70	40	80	55	10
30	60	70	20	70	100	80	45	50
31	95	90	45	70	50	50	45	55
32	50	80	45	60	40	50	40	25
33	70	80	40	70	60	40	25	15
34	50	80	35	70	75	60	60	50
35	55	70	50	85	95	60	50	75
36	45	70	25	70	75	30	50	40
37	60	65	35	50	60	50	45	54
38	50	80	50	80	40	45	55	45
39	70	90	40	60	50	40	40	10
40	55	80	55	80	40	70	75	40
41	75	100	35	70	60	40	55	65
42	40	40	40	50	52	80	25	35
43	60	60	80	50	70	65	70	20

Correlações entre Disciplinas da mesma área			
Port x Lit	0,734	His x Geo	0,624
Mat x Fís	0,037	Bio x Quí	0,647

Outras Correlações (3º Bimestre)			
Port x Mat	0,330	Mat x His	0,423
Port x Fís	-0,009	Mat x Geo	0,765
Port x His	0,422	Mat x Bio	0,714
Port x Geo	0,313	Mat x Quí	0,394
Port x Bio	0,489	Fís x His	0,028
Port x Quí	0,476	Fís x Geo	0,017
Lit x Mat	0,545	Fís x Bio	-0,034
Lit x Fís	0,110	Fís x Quí	0,008
Lit x His	0,376	His x Bio	0,527
Lit x Geo	0,479	His x Quí	0,538
Lit x Bio	0,650	Geo x Bio	0,657
Lit x Quí	0,461	Geo x Quí	0,471

Notas e correlações do 4º Bimestre

Notas do 4º Bimestre (Turma 2011)								
Aluno	Port	Lit	Mat	Fís	His	Geo	Bio	Quí
1	50	50	55	75	75	80	80	75
2	55	60	40	60	60	80	35	75
3	60	80	50	70	85	80	60	50
4	65	70	45	70	80	95	65	65
5	65	60	45	80	75	90	60	70
6	55	60	40	60	80	70	60	85
7	60	50	20	70	70	90	70	65
8	35	50	40	90	70	90	65	65
9	50	50	75	80	80	85	85	50
10	60	50	55	70	60	80	45	70
11	40	50	25	75	50	80	40	75
12	60	100	50	65	75	90	50	70
13	55	100	45	65	90	80	60	35
14	65	50	80	90	80	80	85	75
15	70	50	50	90	80	80	60	70
16	45	50	45	80	70	75	65	45
17	65	50	60	70	80	95	65	80
18	35	50	45	70	80	80	80	55
19	50	50	40	70	85	85	55	80
20	65	60	55	90	80	75	55	80
21	75	60	15	65	90	85	50	60
22	65	50	75	100	85	90	80	95

23	70	50	15	60	80	75	25	40
24	65	50	40	70	80	65	40	70
25	65	50	15	50	75	95	35	85
26	0	0	0	75	0	0	0	0
27	0	0	0	80	0	0	0	0
28	65	50	40	80	70	95	60	60
29	85	70	0	70	80	60	0	55
30	55	50	15	70	70	90	50	85
31	70	60	15	65	70	80	45	80
32	0	50	0	65	70	0	0	0
33	65	50	20	60	70	80	50	40
34	65	50	45	65	75	75	55	95
35	60	50	60	60	80	75	70	70
36	55	50	35	60	75	85	45	30
37	50	60	35	60	90	80	50	80
38	0	50	0	60	70	45	25	0
39	70	50	35	60	75	85	40	45
40	70	50	50	60	60	95	15	85
41	60	50	55	65	65	75	65	65
42	70	70	55	65	90	90	55	65
43	70	30	80	70	70	100	70	56

Correlações entre Disciplinas da mesma área			
Port x Lit	0,485	His x Geo	0,660
Mat x Fís	0,372	Bio x Quí	0,517

Outras Correlações (4º Bimestre)			
Port x Mat	0,441	Mat x His	0,417
Port x Fís	-0,061	Mat x Geo	0,581
Port x His	0,619	Mat x Bio	0,775
Port x Geo	0,781	Mat x Quí	0,518
Port x Bio	0,392	Fís x His	-0,096
Port x Quí	0,693	Fís x Geo	-0,016
Lit x Mat	0,230	Fís x Bio	0,349
Lit x Fís	-0,194	Fís x Quí	0,108
Lit x His	0,731	His x Bio	0,553
Lit x Geo	0,509	His x Quí	0,483
Lit x Bio	0,290	Geo x Bio	0,673
Lit x Quí	0,339	Geo x Quí	0,731

Média anual e respectivas correlações

Média Anual dos alunos (Turma 2011)								
Aluno	Port	Lit	Mat	Fís	His	Geo	Bio	Quí
1	62	71	40	68	62	61	58	60
2	65	78	38	60	55	63	48	60
3	67	88	61	62	72	73	65	71
4	62	63	51	65	68	76	57	62
5	60	65	50	76	67	75	51	72
6	55	65	43	57	65	56	46	60
7	61	75	42	67	61	72	57	60
8	50	62	40	80	62	65	51	52
9	65	78	68	75	66	82	62	70
10	61	72	42	67	55	55	35	60
11	46	67	26	67	60	58	36	60
12	51	83	63	68	60	72	46	61
13	55	62	31	66	60	70	60	51
14	61	63	60	82	62	72	60	63
15	70	60	38	82	70	62	47	60
16	53	43	27	77	50	60	42	35
17	66	78	67	67	66	73	65	67
18	51	73	42	67	67	65	57	63
19	51	62	38	62	63	60	58	67
20	63	78	68	70	75	70	62	65
21	61	73	41	72	67	66	47	43
22	76	80	77	75	77	88	80	91
23	61	68	28	71	65	55	35	48
24	63	68	33	63	58	55	37	60
25	61	68	33	61	63	65	32	60
26	57	43	45	66	60	62	20	36
27	52	46	62	73	55	82	21	42
28	63	77	41	70	63	70	50	60
29	60	81	45	65	62	61	45	46
30	63	70	25	65	70	80	37	60
31	75	80	27	63	55	62	51	60
32	40	70	30	61	56	37	16	45
33	62	73	36	62	62	56	30	40
34	61	72	41	63	67	56	45	60
35	57	72	55	65	73	65	60	64
36	53	60	30	50	70	62	47	40
37	55	71	36	55	67	52	41	60
38	60	76	32	65	51	52	48	53

39	66	70	35	52	59	67	32	41
40	62	80	46	62	52	73	48	60
41	62	80	41	63	56	55	48	60
42	50	50	51	56	66	72	52	63
43	62	52	80	52	57	74	73	40

Correlações entre Disciplinas da mesma área			
Port x Lit	0,419	His x Geo	0,361
Mat x Fís	0,170	Bio x Quí	0,594

Outras Correlações (Média Anual)			
Port x Mat	0,256	Mat x His	0,341
Port x Fís	0,129	Mat x Geo	0,641
Port x His	0,222	Mat x Bio	0,583
Port x Geo	0,383	Mat x Quí	0,388
Port x Bio	0,404	Fís x His	0,105
Port x Quí	0,365	Fís x Geo	0,280
Lit x Mat	0,103	Fís x Bio	0,140
Lit x Fís	-0,055	Fís x Quí	0,231
Lit x His	0,188	His x Bio	0,427
Lit x Geo	0,004	His x Quí	0,495
Lit x Bio	0,298	Geo x Bio	0,523
Lit x Quí	0,512	Geo x Quí	0,387

Após realizar uma análise detalhada, em que foram feitas todas as combinações de pares de disciplinas possíveis e calculado o coeficiente de correlação entre as notas obtidas em todas as épocas do ano, inclusive a média final nas disciplinas, constatou-se que os valores dos coeficientes obtidos para o mesmo par de disciplinas em diferentes épocas do ano não se mantiveram constante, não são valores sempre próximos entre si nem apresentam uma regularidade que nos possibilite afirmar a existência de correlação linear.

Tendo em vista a falta de regularidade apresentada nos resultados, conclui-se que, para as notas estudadas, de modo geral, não se pode afirmar que exista correlação linear, positiva nem negativa, entre as disciplinas. É importante perceber que a conclusão acima não afirma que a correlação linear entre as disciplinas não exista, mas que não temos evidências bastantes para sustentar uma conclusão de que esta correlação exista. As correlações observadas em

alguns testes realizados não podem ser atribuídas exclusivamente às disciplinas, dado que ocorreram ocasionalmente e, na maioria das vezes, de forma aleatória, entre os conjuntos de notas, independente das disciplinas. Além disso, acredita-se na possibilidade da existência de alguma variável oculta (variável que afeta as variáveis em estudo, mas não está incluída nele) como, por exemplo, a maior dedicação de um aluno a uma determinada disciplina tendo em vista a sua necessidade de obter uma nota maior para a sua aprovação.

Capítulo 6

Conclusão

No decorrer do nosso trabalho procuramos mostrar a importância de uma análise mais detalhada no resultado das avaliações. Apresentamos algumas ferramentas que muito auxiliam nesse sentido e as respectivas formas de utilização.

Tivemos o cuidado de apresentar os conteúdos de forma simples e prática para que se tornasse acessível aos profissionais da educação de todas as áreas de conhecimento.

Ressaltamos que as tomadas de decisões, as medidas administrativas e as medidas pedagógicas necessárias para tornar melhor o aproveitamento acadêmico não são contempladas nos objetivos desse trabalho. Essas decisões deverão ser tomadas por cada profissional ou pela equipe pedagógica, segundo suas responsabilidades, suas experiências e de acordo com as peculiaridades de cada situação.

Concluimos que lançar mão de apenas uma medida de tendência central - geralmente a média - é uma forma pouco eficiente na análise de um conjunto de dados, pois esta medida não revela características importantes para a interpretação dos dados.

Mostramos também, através do escore padronizado, que para compararmos, de forma coerente, resultados de avaliações com diferentes graus de dificuldades, temos que utilizar ferramentas adequadas.

Utilizando dados de uma turma do ensino médio realizamos uma análise minuciosa no sentido de tentar constatar a existência ou a inexistência de uma correlação (linear) entre as notas das disciplinas tomadas duas a duas. Concluimos que, na amostra estudada, não há evidência suficiente para afirmar a existência de correlação positiva nem negativa, pois ao compararmos os valores obtidos (do coeficiente de correlação de Pearson) com os parâmetros utilizados (valores críticos) não se observou regularidade consistente entre os referidos valores.

Entendemos que este trabalho é de grande utilidade no sentido de auxiliar os professores e coordenadores pedagógicos na análise dos resultados das avaliações de seus

alunos. Isto feito, tais profissionais terão maior embasamento para propor ações que solucionem possíveis problemas e tornem o aprendizado mais democrático no âmbito escolar.

Capítulo 7

Referências bibliográficas

COSTA NETO, Pedro L. de O. Estatística. São Paulo: Edgard Blücher, 1977.

D'HAINAUT, Louis. Conceitos e métodos da estatística (Tradução). 2. ed. Lisboa: Fundação Calouste Gulbenkian, 1975.v.1.

FONSECA, Jairo Simon da; MARTINS, Gilberto de Andrade. Curso de Estatística. 6. Ed. São Paulo: Atlas, 2012.

FUNDAÇÃO GETÚLIO VARGAS. Nota padronizada. Disponível no portal FVG: [cacr.fgv.br/sites/cacr.fgv.br/files/file/Nota-padronizada\(1\).pdf](http://cacr.fgv.br/sites/cacr.fgv.br/files/file/Nota-padronizada(1).pdf). Acesso em 7 de Fevereiro de 2015.

LEITÃO, Francileide M. P. de Sá. Estatística e realidade no cotidiano dos alunos. Mossoró (RN): UFERSA, 2011.

MILONE, Giuseppe. Estatística: Geral e Aplicada. São Paulo: Thomson Learning, 2006.

MOORE, David S. A estatística básica e sua prática (Tradução). Rio de Janeiro: LTC, 2011.

SILVA, Ermes Medeiros da... |et al.|.Estatística 1. São Paulo: Atlas, 1996.

TRIOLA, M.F. Introdução a Estatística (Tradução). Rio de Janeiro: LTC, 1999 e 2006.