

Média Aparada

Mário Leite

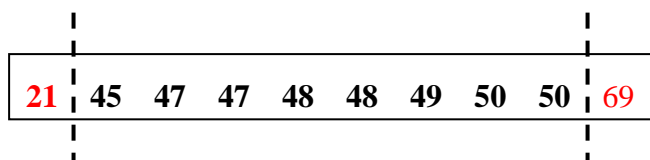
...

Quando obtemos muitos valores e queremos saber qual deles representa todos eles, apelamos para a média; isto é, queremos saber qual valor tende a ser o mais representativo. Em Estatística, a média é definida como “o valor que demonstra a concentração dos dados de uma distribuição, como o ponto de equilíbrio das frequências em um histograma”. Em outras palavras, a média é interpretada como um valor mais provável de uma lista de valores dados. Existem vários tipos de médias; entre elas podemos destacar: média simples, média ponderada, média geométrica, média móvel, média harmônica e média aparada. Todas elas representam o valor mais provável de uma amostra de muitos valores obtidos empiricamente. O primeiro e o último tipo de médias citados acima são bem interessantes, já que o último complementa o primeiro. Por exemplo, é o que acontece em algumas modalidades desportivas, onde a pior marca obtida pelo atleta é excluída da sua série. Vamos considerar dez elementos de uma amostra, como no esquema abaixo.

47	48	45	21	50	48	47	69	49	50
----	----	----	----	----	----	----	----	----	----

A **média simples** é: $(47+48+45+21+50+48+47+69+49+50)/10 = 47.4$

Mas, observe que os valores **21** e **69** são bem discrepantes (*outliers*) dos outros oito valores da amostra. Então, por uma questão de se conseguir uma média mais bem representativa estes dois valores devem ser eliminados (aparados) do conjunto. Na prática isto pode acontecer, pois pode ter havido alguma falha no processo de medição destes dois valores. Deste modo, apenas os outros oito elementos do conjunto é que deverão ser considerados na estatística; e neste caso dizemos que a média obtida é uma “média aparada”; uma média mais representativa do valor central. Ordenando os elementos em ordem crescente, teremos:



No esquema acima a ideia é “cortar”, aparar, os extremos da lista de modo que estes valores “anormais” sejam eliminados, passando a considerar apenas aqueles que estão mais conformes na amostra. Deste modo, os elementos com valores **21** e **69** serão excluídos dos cálculos para não “comprometer” os resultados estatísticos. Agora, recalculando a nova média (*média aparada*), teremos:

$$\text{média aparada} = (45 + 47 + 47 + 48 + 48 + 49 + 50 + 50)/8 = 48.0$$

Note que a média simples (normal) e a média aparada se revelaram diferentes; mas aí surge um outro problema: *como saber quais valores extremos deverão ser eliminados da amostra?*! Por exemplo, se tivesse mais um elemento de valor **22** e mais outro de valor **70** na amostra acima, quais valores deveriam ser excluídos? Nestas circunstâncias o que se faz é considerar um percentual de valores extremos da amostra, que pela literatura estatística pode variar de 2.5% a 5.0%, de acordo com o tamanho do conjunto ou mesmo do interesse do pesquisador. Por exemplo, se uma amostra tem 100 elementos, então a teoria diz que para se obter a “média aparada” a 10% deve-se eliminar os 10 elementos extremos; então deverão ser retirados da amostra os cinco menores valores e também os cinco maiores. A ideia de **média aparada** (também conhecida como *média interna*; do inglês *trimmed-mean*) é uma mistura de *média simples* com *mediana*, criando uma média que não seja influenciada por valores “anormais” do conjunto. Na prática a ideia é “aparar as pontas” de uma amostra ordenada, considerando apenas os “valores do meio” - internos.

E embora com este procedimento a média possa ficar meio “acochambrada” é o que se faz, por exemplo, para apurar a média de inflação ou a média de indicadores sociais sem os dados “estranhos” de uma amostra, quando comparados com os demais valores obtidos na pesquisa. O programa “CalculaMedias”, codificado em Visualg, é uma solução para uma amostra de cinquenta elementos mostrados abaixo, e exibindo: *média normal*, *média aparada*, *variância* e *desvio padrão* (os dois últimos considerando a média aparada).

1, 2, 9, 11, 6, 8, 10, 8, 12, 24, 25, 8, 9, 7, 11, 12, 13, 10, 8, 9, 7, 12, 12, 10,
9, 11, 7, 8, 7, 8, 9, 10, 11, 8, 7, 11, 12, 6, 10, 9, 7, 8, 10, 6, 7, 12, 8, 9, 10, 9

Para o conjunto mostrado acima fica evidente que os valores 1, 2, 24 e 25 podem ser considerados muito discrepantes dos demais; então estes quatro valores deverão ser retirados da amostra, o que significa uma média aparada a 8% (dois valores inferiores e dois valores superiores de uma amostra de tamanho 50).

Algoritmo "CalculaMedias"

```
//Exibe Média Normal, Média Aparada, Variância e Desvio Padrão da amostra.
//Em Visualg
//Autor: Mário Leite
//-----
Const TamMax=100 //limita o tamanho da amostra (poderia ser qualquer um > 2)
Var SomaElem, SomaQuad, MedApa, Variancia, DesvPad, Desvio: real
    Aux, Soma, Media: real
    VetElem: vetor[1..100] de real
    col, Vd, i, j, k, TamAmostra: inteiro
    Resp, Desvc: caractere
Inicio
    LimpaTela
    Soma <- 0.00
    SomaElem <- 0.00
    SomaQuad <- 0.00
    TamAmostra <- 1
    Enquanto ((TamAmostra<2) ou (TamAmostra> TamMax)) Faca
        Escreva("Digite o número de elementos da amostra [2 a 100]: ")
        Leia(TamAmostra)
    FimEnquanto
    Escreval("")
    {Leitura dos valores da amostra}
    Para j De 1 Ate TamAmostra Faca
        Escreva("Digite o elemento", j, " da amostra: ")
        Leia(VetElem[j])
        Soma <- Soma + VetElem[j]
    FimPara //fim da leitura da amostra
    Media <- Soma/TamAmostra
    Media <- Int(Media*10^2+0.50)/10^2 //arredonda para duas decimais
    {Ordena o conjunto em ordem crescente pelo "Método da Bolha"}
    Para i De 1 Ate (TamAmostra-1) Faca
        Para j De (i+1) Ate TamAmostra Faca
            Se(VetElem[i]>VetElem[j]) Entao
                Aux <- VetElem[i]
                VetElem[i] <- VetElem[j]
                VetElem[j] <- Aux
            FimSe
        FimPara
    FimPara //fim da ordenação
    LimpaTela
```

```

{Exibe o conjunto original: com todos os valores da amostra}
Escreval ("Elementos do conjunto original com todos os valores")
col <- 0 //variável de controle de elementos por linha
Para j De 1 Ate (TamAmostra) Faca
    col <- col + 1
    Se (j=TamAmostra) Entao
        Escreva (VetElem[j])
    Senao
        Escreva (VetElem[j], " -")
    FimSe
    Se (col>15) Entao
        Escreval ("" )
        col <- 0
    FimSe
FimPara
Escreval ("" )
Escreval ("" )
{Define e retira da amostra os valores discrepantes}
Escreva ("Quantos valores discrepantes deverão ser excluídos? ")
Leia (Vd)
Escreval ("" )
{Refaz o conjunto sem os Vd valores extremos}
Escreval ("Elementos do conjunto sem os valores extremos")
col <- 0
Para j De 1 Ate (TamAmostra-Vd) Faca
    col <- col + 1
    k <- j + Int (Vd/2)
    VetElem[j] <- VetElem[k]
    {Formata a exibição dos k elementos sem os Vd extremos}
    Se (j=(TamAmostra-Vd)) Entao
        Escreva (VetElem[j])
    Senao
        Escreva (VetElem[j], " -")
    FimSe
    Se (col>15) Entao
        Escreval ("" ) //salta linha
        col <- 0
    FimSe
FimPara
Escreval ("" )
{Calcula Média Aparada, Variância e Desvio Padrão da amostra}
Para j De 1 Ate (TamAmostra-Vd) Faca
    SomaElem <- SomaElem + VetElem[j]
FimPara
MedApa <- SomaElem/(TamAmostra-Vd)
MedApa <- Int (MedApa*10^2+0.50)/10^2 //arredonda para duas decimais
Para j De 1 Ate (TamAmostra-Vd) Faca
    SomaQuad <- SomaQuad + (VetElem[j]-MedApa)*(VetElem[j]-MedApa)
FimPara
Variancia <- SomaQuad/(TamAmostra-Vd)
Variancia <- Int (Variancia*10^2+0.50)/10^2
DesvPad <- RaizQ (variancia)
DesvPad <- Int (DesvPad*10^2+0.50)/10^2
Escreval ("" )
{Exibe os resultados}
Escreval ("Média normal do conjunto: ", Media)
Escreval ("Média aparada do conjunto: ", MedApa)
Escreval ("Variância com média aparada: ", Variancia)
Escreval ("Desvio padrão com média aparada: ", DesvPad)
Escreval ("" )

```

FimAlgoritmo //fim do programa "CalculaMedias"

```
C:\> Console simulando o modo texto do MS-DOS

Elementos do conjunto original com todos os valores
1 - 2 - 6 - 6 - 6 - 7 - 7 - 7 - 7 - 7 - 7 - 7 - 8 - 8 - 8 - 8 -
8 - 8 - 8 - 8 - 8 - 9 - 9 - 9 - 9 - 9 - 9 - 9 - 9 - 10 - 10 - 10 -
10 - 10 - 10 - 10 - 11 - 11 - 11 - 11 - 11 - 12 - 12 - 12 - 12 - 12 - 12 - 13 -
24 - 25

Quantos valores discrepantes deverão se excluídos? 4

Elementos do conjunto sem os valores extremos
6 - 6 - 6 - 7 - 7 - 7 - 7 - 7 - 7 - 7 - 8 - 8 - 8 - 8 - 8 - 8 -
8 - 8 - 8 - 9 - 9 - 9 - 9 - 9 - 9 - 9 - 9 - 10 - 10 - 10 - 10 - 10 -
10 - 10 - 11 - 11 - 11 - 11 - 11 - 12 - 12 - 12 - 12 - 12 - 12 - 13

Média normal do conjunto:  9.46
Média aparada do conjunto:  9.15
Variância com média aparada:  3.48
Desvio padrão com média aparada:  1.87

>>> Fim da execução do programa !
```

Figura 1 - Entradas e saídas do programa “CalculaMedias”