

**IPECE**

**Textos  
para Discussão**

Nº 139 – abril/2026

## **Considerações sobre a forma de cálculo do IDH**

**Autor: José Fabio Bezerra Montenegro**

**Colaborador: Aprígio Botelho Lócio**



## Governador do Estado do Ceará

Elmano de Freitas da Costa

## Vice-Governadora do Estado do Ceará

Jade Afonso Romero

## Secretaria do Planejamento e Gestão – SEPLAG

Alexandre Sobreira Cialdini - Secretário

Caio Hugo Carvalho Vitor - Secretário Executivo de Gestão de Compras e Patrimônio

Naiana Corrêa Lima Peixoto - Secretária Executiva de Planejamento e Orçamento

José Garrido Braga Neto - Secretário Executivo de Gestão de Pessoas

Daniel de Carvalho Bentes - Secretário Executivo de Modernização e Governo Digital

Francisca Rejane Araujo Felipe Pessoa de Albuquerque - Secretária Executiva de Planejamento e Gestão Interna

## Instituto de Pesquisa e Estratégia Econômica do Ceará – IPECE

### Diretor Geral

Alfredo José Pessoa de Oliveira

### Diretoria de Estudos Econômicos – DIEC

Ricardo Antônio de Castro Pereira

### Diretoria de Estudos Sociais – DISOC

José Meneleu Neto

### Diretoria de Estudos de Gestão Pública – DIGEP

José Fábio Bezerra Montenegro

### Gerência de Estatística, Geografia e Informações – GEGIN

Rafaela Martins Leite Monteiro

## Texto para Discussão – Nº 139 – abril de 2026

### Unidade Responsável:

Diretoria de Estudos de Gestão Pública – DIGEP

### Elaboração:

José Fábio Bezerra Montenegro (Diretor)

### Colaboração:

Aprígio Botelho Lócio (Assistente de Gestão)

O Instituto de Pesquisa e Estratégia Econômica do Ceará (IPECE) é uma autarquia vinculada à Secretaria do Planejamento e Gestão do Estado do Ceará. Fundado em 14 de abril de 2003, o IPECE é o órgão do Governo responsável pela geração de estudos, pesquisas e informações socioeconômicas e geográficas que permitem a avaliação de programas e a elaboração de estratégias e políticas públicas para o desenvolvimento do Estado do Ceará.

**Missão:** Gerar e disseminar conhecimento e informações, subsidiar a formulação e avaliação de políticas públicas e assessorar o Governo nas decisões estratégicas, contribuindo para o desenvolvimento sustentável do Ceará.

**Visão:** Até 2031, consolidar-se como referência em inteligência pública e assessoramento estratégico ao Governo do Ceará, ampliando sua capacidade de produzir e disseminar conhecimento qualificado, inovador e orientado às políticas públicas efetivas e sustentáveis.

**Valores:** ética, transparência e impessoalidade; competência, comprometimento e senso de equipe; compromisso com a sociedade e valorização do ser humano; autonomia técnica; e rigor científico e inovação.

Instituto de Pesquisa e Estratégia Econômica do Ceará (IPECE)  
Av. Gal. Afonso Albuquerque Lima, s/n | Edifício SEPLAG | Térreo  
Cambéba | Cep: 60.822-325 |  
Fortaleza, Ceará, Brasil | Telefone: (85) 2018-2639  
[www.ipece.ce.gov.br](http://www.ipece.ce.gov.br)

## Sobre o Texto para Discussão

A Série **Textos para Discussão** do Instituto de Pesquisa e Estratégia Econômica do Ceará (IPECE) tem como objetivo a divulgação de estudos elaborados ou coordenados por servidores do órgão, que possam contribuir para a discussão de temas de interesse do Estado. As conclusões, metodologia aplicada ou propostas contidas nos textos são de inteira responsabilidade do(s) autor(es) e não exprimem, necessariamente, o ponto de vista ou o endosso do Instituto de Pesquisa e Estratégia Econômica do Ceará - IPECE, da Secretaria de Planejamento e Gestão ou do Governo do Estado do Ceará.

Instituto de Pesquisa e Estratégia Econômica do Ceará – IPECE 2026

Texto para discussão / Instituto de Pesquisa e Estratégia Econômica do Ceará (IPECE) / Fortaleza – Ceará: IPECE, 2026

ISSN: 1983-4969

1. Economia Brasileira. 2. Economia Cearense. 3. Aspectos Econômicos. 4. Aspectos Sociais. 5. Mercado de Trabalho. 6. Finanças Públicas. 7. Gestão Pública.

## Nesta Edição

O presente estudo tem como objetivo a avaliação das justificativas da escolha da média geométrica – em detrimento da aritmética - no cálculo do Índice de Desenvolvimento Humano (IDH), que considera os indicadores de saúde, educação e renda. Vale notar que a média aritmética ainda é aplicada ao Coeficiente de Desigualdade Humana (CDH).

## 1. Introdução

O objetivo aqui é avaliar as justificativas da escolha da média geométrica – em detrimento da aritmética - no cálculo do Índice de Desenvolvimento Humano (IDH), que considera os indicadores de saúde, educação e renda. Vale notar que a média aritmética ainda é aplicada ao Coeficiente de Desigualdade Humana (CDH).

Para essa análise, serão utilizadas notações matemáticas, fórmulas e elementos de cálculo diferencial de várias variáveis. No entanto, o texto foi estruturado para ser autossuficiente, apresentando todas as definições e argumentações clássicas; para facilitar a leitura, as demonstrações detalhadas foram apresentadas no apêndice.

Embora os resultados aqui expostos sejam válidos para qualquer conjunto finito de números positivos, serão focados em conjuntos de três valores entre zero e um. Ainda que a notação vetorial fosse rigorosamente mais adequada por preservar a ordem dos indicadores (saúde, educação e renda), optou-se pela notação de conjunto, uma vez que a ordem não altera as argumentações e conclusões.

## 2. Notações e definições

Considere o conjunto

$$x = \{x_1, x_2, x_3\} \subset [0,1].$$

A média aritmética, ou simplesmente *média*, dos dados em  $x$  é definida por

$$\bar{x} = \frac{x_1 + x_2 + x_3}{3}$$

A média *geométrica* dos dados em  $x$  é definida por

$$\bar{x}_g = \sqrt[3]{x_1 x_2 x_3}$$

Considere, também, a *variância*, que é um tipo de dispersão, dos dados em  $x$ , definida por

$$\sigma_x^2 = \frac{1}{3} ((x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2)$$

e uma segunda *medida de dispersão*, definida por

$$\bar{\sigma}_x^2 = (x_1 - x_2)^2 + (x_1 - x_3)^2 + (x_2 - x_3)^2$$

Esta última diz o quão próximo os dados estão entre si e a variância diz o quão próximo os dados estão de sua média. Verifica-se uma certa equivalência entre essas duas dispersões, no sentido de que, se uma se aproxima de zero, o mesmo acontece com a outra, devido a seguinte relação entre elas, demonstrada no apêndice.

$$\sigma_x^2 = \frac{1}{9} \bar{\sigma}_x^2 \quad (1)$$

Pode-se dizer que os números  $x_1, x_2$  e  $x_3$  diferem em magnitude por um valor menor que  $\epsilon > 0$  se  $\bar{\sigma}_x^2 < \epsilon$ . Observe que, pela identidade em (1), que isto é equivalente a dizer que a variância entre os dados  $x_1, x_2$  e  $x_3$  é menor que  $\epsilon/9$ .

### 3. Justificativa que não se sustenta

No relatório técnico *Human Development Report 2025 technical notes*<sup>1</sup> há dois comentários que pretendem justificar a escolha da média geométrica para o cálculo do IDH em detrimento da média aritmética. "*When all inequalities in dimensions are of a similar magnitude, the coefficient of human inequality and the loss in HDI value differ negligibly. When inequalities differ in magnitude, the loss in HDI value tends to be higher than the coefficient of human inequality*". "Quando todas as desigualdades nas dimensões são de magnitude semelhantes, o coeficiente de desigualdade humana e a perda no valor do IDH diferem insignificamente. Quando as desigualdades diferem em magnitude, a perda no valor do IDH tende a ser maior do que o coeficiente de desigualdade humana".

A partir dessas duas afirmações, mas procurando argumentar de forma mais rigorosa. A primeira afirmação é sobre a diferença entre a média e a média geométrica,  $\bar{x} - \bar{x}_g$ , que é pequena, "diferem insignificamente", se  $x_1, x_2$  e  $x_3$  diferem em magnitude por um valor pequeno, "são de magnitude semelhantes". Reescrevendo com mais rigor, trocando a expressão "dimensões são de magnitudes semelhantes" por dispersão entre as dimensões próximas de zero e "desigualdade humana e a perda no valor do IDH diferem insignificamente" pela diferença entre desigualdade humana e valor do IDH próximos de zero, tem-se a seguinte afirmação.

**Afirmação 1:** A diferença  $\bar{x} - \bar{x}_g$  se aproxima de zero quando a dispersão se aproxima de zero, ou ainda, na notação de limite

$$\lim_{\sigma_x^2 \rightarrow 0} (\bar{x} - \bar{x}_g) = 0$$

Tal afirmação é verdadeira e sua demonstração será apresentada no apêndice desta nota.

Já na segunda afirmação é dito que a "perda no valor do IDH tende a ser maior do que a perda do coeficiente de desigualdade humana, quando as desigualdades diferem em magnitude", isto é, quando a variância cresce ( $\bar{\sigma}_x^2 < \bar{\sigma}_y^2$ ), então existe perda do IDH ( $\bar{x}_g > \bar{y}_g$ ), e tal perda é maior que a perda da média, ( $\bar{x}_g - \bar{y}_g > \bar{x} - \bar{y} > 0$ ). A ideia é que uma maior dispersão penalize mais o IDH que o CDH.

**Afirmação 2:** Sejam  $x = \{x_1, x_2, x_3\}$  e  $y = \{y_1, y_2, y_3\}$  subconjuntos do intervalo  $[0,1]$ . Se  $\sigma_x^2 < \sigma_y^2$ , então

$$\bar{x}_g - \bar{y}_g > \bar{x} - \bar{y} > 0 \quad (2)$$

<sup>1</sup> Relatório técnico que detalha a metodologia, as fontes de dados e os cálculos utilizados para compor o Índice de Desenvolvimento Humano (IDH) e outros índices compostos do relatório anual. Publicado pelo *Human Development Report Office (HDRO)*, que faz parte do Programa das Nações Unidas para o Desenvolvimento (PNUD) Disponível em: [https://hdr.undp.org/sites/default/files/2025\\_HDR/HDR25\\_Technical\\_Notes.pdf](https://hdr.undp.org/sites/default/files/2025_HDR/HDR25_Technical_Notes.pdf)

O problema é que tal afirmação não é verdadeira. De fato, pode ocorrer da média decrescer ( $\bar{x} > \bar{y}$ ) e a variância crescer ( $\sigma_x^2 < \sigma_y^2$ ), mas a média geométrica crescer ( $\bar{x}_g < \bar{y}_g$ ). Portanto temos,  $\bar{x}_g - \bar{y}_g < 0$  e a primeira desigualdade em (2) não se verifica. A afirmação abaixo diz que existem uma infinidade de exemplos desse tipo.

**Afirmação 3:** Para todo  $x = \{x_1, x_2, x_3\}$ , com  $x_i \neq x_j$  para todo  $i \neq j$ , existe  $y = \{y_1, y_2, y_3\}$  com

$$\sigma_x^2 < \sigma_y^2, \quad \bar{x} > \bar{y} \quad \text{e} \quad \bar{x}_g < \bar{y}_g. \quad (3)$$

A demonstração desta afirmação será apresentada no apêndice.

Como as funções  $\bar{x}$ ,  $\bar{x}_g$  e  $\sigma_x^2$  são contínuas, pode-se concluir que a probabilidade de ocorrer as desigualdades em (3) é positiva. Será calculada tal probabilidade na seção 3.

Uma identidade importante, para compreender melhor a média geométrica, é apresentada abaixo e demonstrada do apêndice.

**Lema 1:** Seja  $x = \{x_1, x_2, x_3\}$  um conjunto de três números positivos. Então,

$$\bar{x}_g^3 = \bar{x}^3 - \frac{3}{2}\bar{x}\sigma_x^2 + (x_1 - \bar{x})(x_2 - \bar{x})(x_3 - \bar{x}) \quad (4)$$

Em particular, se  $y = \{y_1, y_2, y_3\}$  é um outro conjunto de dados tal que  $\bar{x} = \bar{y}$  e  $\sigma_x^2 = \sigma_y^2$ , então

$$\bar{x}_g^3 - \bar{y}_g^3 = (x_1 - \bar{x})(x_2 - \bar{x})(x_3 - \bar{x}) - (y_1 - \bar{x})(y_2 - \bar{x})(y_3 - \bar{x})$$

A identidade em (4) pode ser obtida por expansão em série de Taylor da função  $\bar{x}_g^3$  em torno do ponto  $(\bar{x}, \bar{x}, \bar{x})$ , ou ser verificada, simplesmente expandindo o lado direito da igualdade.

#### 4. Uma outra média a ser considerada

Um dos problemas da média aritmética é que esta não "enxerga" a variância dos dados. Pode-se ter dois conjuntos  $x = \{0.50, 0.40, 0.60\}$  e  $y = \{0.20, 0.50, 0.80\}$  com mesma média aritmética ( $\bar{x} = \bar{y} = 0.50$ ) e variâncias distintas ( $\sigma_x^2 < \sigma_y^2$ ). Se considerar  $x$  e  $y$  como dois países e seus elementos como os indicadores de saúde, educação e renda, nessa ordem, percebe-se grandes diferenças entre eles,  $y$  possui um indicador de renda bem maior, possivelmente concentrada, pois o indicador de saúde de sua população é bem menor que de  $x$ . Então, se quiser construir um índice que leve em consideração também a variância dos indicadores das três dimensões, este índice não pode ser obtido pela média dos indicadores, como no caso do Coeficiente de Desigualdade Humana (CDH).

O problema da escolha da média geométrica para o cálculo do IDH é que esta não pode ser explicada somente pela média e a variância, logo não se pode utilizar estas medidas para justificar a escolha da média geométrica para o cálculo do IDH em detrimento da média aritmética. Os dois conjuntos de dados  $x = \{0.52, 0.52, 0.55\}$  e  $y = \{0.51, 0.54, 0.54\}$  possuem a mesma média ( $\bar{x} = \bar{y}$ ) e a mesma variância ( $\sigma_x^2 = \sigma_y^2$ ), porém médias geométricas distintas ( $\bar{x}_g > \bar{y}_g$ ). Ou seja, existe algo além da média e da variância que determinam o IDH.

Tal fato pode ser também observado, de forma mais analítica, em (4). Além disso a variação da média geométrica não se explica somente pelo crescimento da média e o decréscimo da variância

dos dados. De fato, pode ocorrer da média crescer ( $\bar{x} < \bar{y}$ ) e a variância decrescer ( $\sigma_x^2 > \sigma_y^2$ ), mas a média geométrica decrescer ( $x_g > y_g$ ). Ou seja, pode ocorrer que, de um ano para o outro, o Brasil melhorar seu CDH, diminuir a variância entre os indicadores de saúde, educação e renda, mas, mesmo com esse melhor desempenho, piorar seu IDH. É possível ver que a probabilidade de tal fato ocorrer é pequena, mas existe, é positiva.

Por esse motivo propõe-se, para o cálculo do IDH, a *média de produtos binários*, que não possui tal deficiência, pois possui uma relação simples com a média e a dispersão, seu quadrado é uma combinação linear entre o quadrado da média e a variância. Tal média é a raiz quadrada da média aritmética dos produtos de todos os pares possíveis entre os termos, ou seja,

$$\hat{x} = \sqrt{\frac{x_1x_2 + x_1x_3 + x_2x_3}{3}} \quad (5)$$

e satisfaz a identidade abaixo, que pode também ser usada para defini-la.

**Lema 2:** Para todo  $x = \{x_1, x_2, x_3\} \subset [0,1]$  vale

$$\hat{x}^2 = \bar{x}^2 - \frac{1}{2}\sigma_x^2 \quad (6)$$

A partir desse lema pode-se garantir que se  $\bar{x} = \bar{y}$  e  $\sigma_x^2 = \sigma_y^2$ , então  $\hat{x} = \hat{y}$ . A proposição abaixo, corresponde a afirmação 1 para a média geométrica.

**Proposição 1:**

$$\lim_{\sigma_x^2 \rightarrow 0} (\bar{x} - \hat{x}) = 0.$$

Uma demonstração desta proposição é apresentada no apêndice. Diferente da média geométrica, tal média satisfaz a seguinte proposição.

**Proposição 2:** Sejam  $x = \{x_1, x_2, x_3\}$  e  $y = \{y_1, y_2, y_3\}$ , com  $x_i, y_i \in (0,1)$  para todo  $i = 1,2,3$ . Se  $\bar{x} \geq \bar{y}$  e  $\sigma_x^2 \leq \sigma_y^2$ , então  $\hat{x} \geq \hat{y}$ .

Prova: Pelas hipóteses e por (6),

$$\hat{x}^2 = \bar{x}^2 - \frac{1}{2}\sigma_x^2 \geq \bar{y}^2 - \frac{1}{2}\sigma_y^2 = \hat{y}^2 \Rightarrow \hat{x}^2 \geq \hat{y}^2 \Rightarrow \hat{x} \geq \hat{y}.$$

## 5. Probabilidade de um evento inesperado

É visto na Afirmação 3 que existem uma infinidade de pares de conjuntos de dados,  $x = \{x_1, x_2, x_3\}$  e  $y = \{y_1, y_2, y_3\}$ , tais que

$$\bar{x} \leq \bar{y}, \quad \bar{\sigma}_x^2 \geq \bar{\sigma}_y^2 \quad \text{e} \quad \bar{x}_g > \bar{y}_g. \quad (7)$$

Ou, ainda

$$\bar{x} \geq \bar{y}, \quad \bar{\sigma}_x^2 \leq \bar{\sigma}_y^2 \quad \text{e} \quad \bar{x}_g < \bar{y}_g.$$

O objetivo agora é calcular a probabilidade desses eventos inesperados ocorrerem. Como esses dados estão todos entre 0 e 1, o espaço amostral é o cubo  $[0,1]^6$  em  $\mathbb{R}^6$  e a probabilidade que nos interessa é o volume do conjunto

$$\{(x, y) \in R^6; \bar{x} \leq \bar{y}, \bar{\sigma}_x^2 \geq \bar{\sigma}_y^2 \text{ e } \bar{x}_g > \bar{y}_g\} \cup \{(x, y) \in R^6; \bar{x} \geq \bar{y}, \bar{\sigma}_x^2 \leq \bar{\sigma}_y^2 \text{ e } \bar{x}_g < \bar{y}_g\},$$

pois o cubo  $[0,1]^6$  tem volume 1. Verifica-se que tal conjunto é a união de dois conjuntos disjuntos de mesmo volume, pois se  $(x, y)$  pertence ao primeiro conjunto, então  $(y, x)$  pertence ao segundo, de forma que, tal probabilidade será duas vezes o volume do primeiro conjunto.

Um cálculo para isso é bem complicado e não se sabe que seja possível uma solução analítica. Por esse motivo contorna-se esse problema considerando um espaço amostral finito. Como, na divulgação do IDH, os indicadores, das dimensões de saúde, educação e renda, são apresentados com duas casas decimais, será considerado o espaço amostral como o conjunto de todos vetores  $(x_1, x_2, x_3, y_1, y_2, y_3)$  cujas entradas são números entre 0 e 1, com somente duas casas decimais. Em notação de conjunto, o espaço amostral é

$$\{(j_1/100, j_2/100, j_3/100, j_4/100, j_5/100, j_6/100,); j_i \in N \text{ e } 1 \leq j_i \leq 100, \forall i\}$$

Tal conjunto possui  $100^6$  elementos. Naturalmente será utilizado um computador para testar, para todos os elementos do espaço amostral, se ocorrem as desigualdades em (7), contar o número de todos esses eventos, denotado por  $n$ , e a probabilidade procurada será  $2n/100^6$ . O programa é bem simples.

Primeiro, as desigualdades em (7) ocorrerão se, e somente se,

$$\begin{cases} j_1 + j_2 + j_3 \leq j_4 + j_5 + j_6 \\ (j_1 - j_2)^2 + (j_1 - j_3)^2 + (j_2 - j_3)^2 \geq (j_3 - j_4)^2 + (j_3 - j_5)^2 + (j_4 - j_5)^2 \\ j_1 j_2 j_3 > j_4 j_5 j_6 \end{cases}$$

Na segunda desigualdade utiliza-se a identidade em (1).

No programa, para cada elemento do espaço amostral, deve-se fazer as três perguntas acima, ou seja, serão necessárias  $3 \times 100^6$  cálculos. O programa mais simples utiliza seis laços do tipo *for*, cada um variando de 1 a 100, e no último laço deve-se verificar se as três desigualdades ocorrem. Recomenda-se utilizar a linguagem C++, que leva alguns minutos para calcular tal probabilidade, outras linguagens podem levar algumas horas para fazer esse cálculo.

É apresentado aqui um código fonte simples em C++, não o mais eficiente, para calcular essa probabilidade.

```
int main() {
    // Inicializa contador
    long long cont = 0;
    const int k = 100;
    const long long total_iteracoes =
    (long long)k*k*k*k*k*k;k;
```

```

for (int i1=1; i1 <= k; ++i1) {
  for (int i2=1; i2 <= k; ++i2) {
    for (int i3=1; i3 <= k; ++i3) {
      int Max = i1+i2+i3;
      int Vax = (i1-i2)*(i1-i2)
                + (i1-i3)*(i1-i3)
                + (i2-i3)*(i2-i3);
      long long Mgx = (long long)i1 * i2 * i3;
      for (int i4=1; i4 <= k; ++i4) {
        for (int i5=1; i5 <= k; ++i5) {
          for (int i6=1; i6 <= k; ++i6) {
            int May = i4+i5+i6;
            int Vay = (i4-i5)*(i4-i5)
                      + (i4-i6)*(i4-i6)
                      + (i5-i6)*(i5-i6);
            long long Mgy = (long long)i4 * i5 * i6;
            if ((Max <= May) && (Vax >= Vay)
                && (Mgx > Mgy)) {
              cont++;
            }
          }
        }
      }
    }
  }
}

// Resultado final
double resultado = 2 * 100.0 * cont / total_iteracoes;

// Configura a precisao de saida e imprime resultado
std::cout << std::fixed << std::setprecision(10) <<
resultado << std::endl;

return 0;
}

```

O resultado dessa conta, o cálculo da probabilidade, é de aproximadamente **0,310939%**.

## 6. Conclusão

Parece um esforço irrelevante? Ocorre que, quando se justifica o uso da média geométrica no cálculo do IDH — em vez da aritmética — como uma forma de penalizar desigualdades entre as dimensões (como explica a IA Gemini), deve-se ver tal afirmação não como verdade absoluta, mas uma verdade com margem de erro de 0,31%.

## 7 Apêndice

### 7.1 Relação entre dispersões

**Lema 3** *Seja  $x = \{x_1, x_2, x_3\}$  um conjunto de três números reais. Então*

$$\sigma_x^2 = \frac{1}{3} \bar{\sigma}_x^2$$

**Prova:** De fato, por definição

$$\sigma_x^2 = \frac{1}{3} [(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2]$$

onde  $\bar{x} = (x_1 + x_2 + x_3)/3$ . Logo,

$$\begin{aligned} \sigma_x^2 &= \frac{1}{3} \left[ \left( \frac{2x_1 - x_2 - x_3}{3} \right)^2 + \left( \frac{2x_2 - x_1 - x_3}{3} \right)^2 + \left( \frac{2x_3 - x_1 - x_2}{3} \right)^2 \right] \\ &= \frac{6}{27} (x_1^2 + x_2^2 + x_3^2 - x_1x_2 - x_1x_3 - x_2x_3) \\ &= \frac{1}{9} ((x_1 - x_2)^2 + (x_1 - x_3)^2 + (x_2 - x_3)^2) = \frac{1}{9} \bar{\sigma}_x^2 \end{aligned}$$

### 7.2 Comparações entre as três médias aqui consideradas

**Lema 4** *Seja  $x = \{x_1, x_2, x_3\}$  um conjunto de três números reais. Então,*

$$\bar{x}_g \leq \hat{x} \leq \bar{x}. \quad (8)$$

com igualdade somente quando a dispersão se anula,  $\sigma_x^2 = 0$ .

**Prova:** A segunda desigualdade sai da identidade do Lema 1.

$$\hat{x}^2 = \bar{x}^2 - \frac{1}{2} \sigma_x^2 \leq \bar{x}^2 \Rightarrow \hat{x} \leq \bar{x}$$

com igualdade somente quando  $\sigma_x^2 = 0$ .

A primeira desigualdade é obtida pela desigualdade entre a média geométrica e a média aritmética

$$\sqrt[3]{abc} \leq \frac{a+b+c}{3}, \text{ para } a, b, c > 0.$$

Basta considerar  $a = xy$ ,  $b = xz$  e  $c = yz$  e obter

$$\sqrt[3]{(xy)(xz)(yz)} = (\sqrt[3]{xyz})^2 \leq \frac{xy + xz + yz}{3} \Rightarrow \bar{x}_g \leq \hat{x} \quad \square$$

### 7.3 Prova da Afirmação 1

**Afirmação 1:** A diferença  $\bar{x} - \bar{x}_g$  se aproxima de zero quando a dispersão se aproxima de zero, ou na notação de limite

$$\lim_{\sigma_x^2 \rightarrow 0} (\bar{x} - \bar{x}_g) = 0.$$

**Prova:** Pelo mesmo raciocínio no início da prova da Proposição 1, dado  $\epsilon > 0$ , existe  $r \in (0, 1)$  tal que  $\bar{x} - \bar{x}_g < \epsilon$  para todo  $(x_1, x_2, x_3)$  a uma distância da origem menor que  $r$ ,

$$x_1^2 + x_2^2 + x_3^2 < r^2$$

Passemos a estudar o que acontece na região dos pontos  $(x_1, x_2, x_3)$  para os quais

$$x_1^2 + x_2^2 + x_3^2 \geq r^2.$$

Neste caso, como já vimos em (14),

$$\max\{x_1, x_2, x_3\} \geq \frac{\sqrt{3}}{3} r, \quad \forall i = 1, 2, 3. \quad (9)$$

Usando a identidade

$$a^3 + b^3 + c^3 - 3abc = \frac{1}{2}(a + b + c)[(a - b)^2 + (a - c)^2 + (b - c)^2]$$

com  $a = \sqrt[3]{x_1}$ ,  $b = \sqrt[3]{x_2}$  e  $c = \sqrt[3]{x_3}$ , temos

$$\begin{aligned} x_1 + x_2 + x_3 - 3\sqrt[3]{x_1 x_2 x_3} = \\ \frac{1}{2}(\sqrt[3]{x_1} + \sqrt[3]{x_2} + \sqrt[3]{x_3})[(\sqrt[3]{x_1} - \sqrt[3]{x_2})^2 + (\sqrt[3]{x_1} - \sqrt[3]{x_3})^2 + (\sqrt[3]{x_2} - \sqrt[3]{x_3})^2] \end{aligned}$$

e dividindo a identidade acima por três, temos

$$\begin{aligned} \bar{x} - \bar{x}_g = \\ \frac{1}{6}(\sqrt[3]{x_1} + \sqrt[3]{x_2} + \sqrt[3]{x_3}) [(\sqrt[3]{x_1} - \sqrt[3]{x_2})^2 + (\sqrt[3]{x_1} - \sqrt[3]{x_3})^2 + (\sqrt[3]{x_2} - \sqrt[3]{x_3})^2] \end{aligned} \quad (10)$$

Agora consideramos a identidade  $a^3 - b^3 = (a - b)(a^2 + ab + b^2)$ , para  $a = \sqrt[3]{x_i}$  e  $b = \sqrt[3]{x_j}$  dados acima, e achamos a seguinte identidade

$$x_i - x_j = (\sqrt[3]{x_i} - \sqrt[3]{x_j}) \left( \sqrt[3]{x_i^2} + \sqrt[3]{x_i} \sqrt[3]{x_j} + \sqrt[3]{x_j^2} \right) \quad (11)$$

Sem perda de generalidade podemos supor que  $\max\{x_1, x_2, x_3\} = x_3$ . E só agora vamos usar o fato de que  $0 \leq x_i \leq 1$ , para todo  $i = 1, 2, 3$ . Desta forma, por (9), (10) e (11)

$$\begin{aligned} \bar{x} - \bar{x}_g &\leq \frac{3}{6} [2(\sqrt[3]{x_1} - \sqrt[3]{x_3})^2 + 2(\sqrt[3]{x_2} - \sqrt[3]{x_3})^2] \\ &= \frac{(x_1 - x_3)^2}{\left(\sqrt[3]{x_1^2} + \sqrt[3]{x_1} \sqrt[3]{x_3} + \sqrt[3]{x_3^2}\right)^2} + \frac{(x_2 - x_3)^2}{\left(\sqrt[3]{x_2^2} + \sqrt[3]{x_2} \sqrt[3]{x_3} + \sqrt[3]{x_3^2}\right)^2} \end{aligned}$$

$$\begin{aligned} &\leq \frac{(x_1 - x_3)^2}{\left(\sqrt[3]{x_3^2}\right)^2} + \frac{(x_2 - x_3)^2}{\left(\sqrt[3]{x_3^2}\right)^2} = \frac{1}{\left(\sqrt[3]{x_3}\right)^4} [(x_1 - x_3)^2 + (x_2 - x_3)^2] \\ &\leq \frac{1}{\left(\sqrt[3]{r/\sqrt{3}}\right)^4} [(x_1 - x_3)^2 + (x_1 - x_2)^2 + (x_2 - x_3)^2] \\ &= \frac{3^{4/6}}{r^{4/3}} [(x_1 - x_3)^2 + (x_1 - x_2)^2 + (x_2 - x_3)^2] = \frac{3\sqrt[3]{9}}{r^{4/3}} \sigma_x^2 \end{aligned}$$

Logo, por (1), temos

$$\bar{x} - \bar{x}_g \leq \frac{9^{4/3}}{r^{4/3}} \sigma_x^2 \quad (12)$$

sempre que  $x_1^2 + x_2^2 + x_3^2 \geq r^2$ . Agora consideramos

$$\delta = \frac{r^{4/3} \epsilon}{9^{4/3}}.$$

Desta forma, se  $\sigma_x^2 < \delta$ , então

$$\bar{x} - \bar{x}_g \leq \frac{9^{4/3}}{r^{4/3}} \sigma_x^2 < \frac{9^{4/3}}{r^{4/3}} \delta = \epsilon \Rightarrow \bar{x} - \bar{x}_g < \epsilon$$

sempre que  $x_1^2 + x_2^2 + x_3^2 \geq r^2$  e  $\sigma_x^2 < \delta$  e já vimos que  $\bar{x} - \bar{x}_g < \epsilon$  também ocorre quando  $x_1^2 + x_2^2 + x_3^2 < r^2$ . Ou seja,

$$\lim_{\sigma_x^2 \rightarrow 0} (\bar{x} - \bar{x}_g) = 0.$$

## 7.4 Prova da Proposição 1

**Proposição**

$$\lim_{\sigma_x^2 \rightarrow 0} (\bar{x} - \hat{x}) = 0.$$

**Prova:** Pela definição de limite, dado qualquer número  $\epsilon > 0$ , devemos exibir  $\delta > 0$  tal que

$$\bar{x} - \hat{x} < \epsilon$$

sempre que  $\sigma_x^2 < \delta$ .

Como as funções  $\bar{x}$  e  $\hat{x}$  são contínuas e se anulam na origem,  $(x_1, x_2, x_3) = (0, 0, 0)$ , temos que  $\bar{x}$  e  $\hat{x}$  se aproximam de zero quando  $(x_1, x_2, x_3)$  se aproxima da origem. De forma que existe  $0 < r < 1$  tal que

$$\bar{x}(x_1, x_2, x_3) < \frac{\epsilon}{2} \text{ e } \hat{x}(x_1, x_2, x_3) < \frac{\epsilon}{2}$$

sempre que a distância de  $(x_1, x_2, x_3)$  à origem seja menor que  $r$ , ou seja,

$$x_1^2 + x_2^2 + x_3^2 < r^2.$$

Neste caso,

$$\bar{x} - \hat{x} \leq \bar{x} + \hat{x} < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon \Rightarrow \bar{x} - \hat{x} < \epsilon \quad (13)$$

Passemos a estudar os pontos  $(x_1, x_2, x_3)$  tais que  $x_1^2 + x_2^2 + x_3^2 \geq r^2$ . Para tais números temos que o maior deles,  $\max\{x_1, x_2, x_3\}$ , satisfaz

$$\max\{x_1, x_2, x_3\}^2 \geq \frac{1}{3}r^2 \quad (14)$$

Logo,

$$\bar{x} = \frac{1}{3}(x_1 + x_2 + x_3) \geq \frac{1}{3} \max\{x_1, x_2, x_3\} \geq \frac{r}{3\sqrt{3}} \Rightarrow \bar{x} \geq \frac{\sqrt{3}}{9} r$$

e como  $\hat{x} \geq 0$  temos que

$$\bar{x}^2 - \hat{x}^2 = (\bar{x} - \hat{x})(\bar{x}^2 + \bar{x}\hat{x} + \hat{x}^2) \geq \frac{r^2}{27}(\bar{x} - \hat{x})$$

e, pela identidade do Lema 1,

$$\bar{x} - \hat{x} \leq \frac{27}{r^2}(\bar{x}^2 - \hat{x}^2) = \frac{27\sigma_x^2}{2r^2} \Rightarrow \bar{x} - \hat{x} \leq \frac{27\sigma_x^2}{2r^2}$$

Agora podemos fazer a escolha do  $\delta > 0$ . Considere  $\delta = 2r^2\epsilon/27$ . Logo, se  $(x_1, x_2, x_3)$  satisfaz  $x_1^2 + x_2^2 + x_3^2 \geq r^2$  e  $\sigma_x^2 < \delta$ , então pela desigualdade acima

$$\bar{x} - \hat{x} \leq \frac{27\sigma_x^2}{2r^2} < \frac{27\delta}{2r^2} = \epsilon \Rightarrow \bar{x} - \hat{x} < \epsilon$$

sempre que  $\sigma_x^2 < \delta$  e  $x_1^2 + x_2^2 + x_3^2 \geq r^2$ . Mas  $r > 0$  já foi escolhido de forma que  $\bar{x} - \hat{x} < \epsilon$  sempre que  $x_1^2 + x_2^2 + x_3^2 < r^2$ .

Ou seja, em qualquer dos casos, dado  $\epsilon > 0$ , existe  $\delta = 2r^2\epsilon$ , para algum  $r \in (0, 1)$ , tal que

$$\bar{x} - \hat{x} < \epsilon$$

sempre que  $\sigma_x^2 < \delta$ .

## 7.5 Prova do Lema 1

Iremos expandir, em série de Taylor, a função

$$f(x_1, x_2, x_3) = x_1x_2x_3$$

em torno do ponto  $(\bar{x}, \bar{x}, \bar{x})$ . Temos que, nesse ponto

$$\begin{aligned} \frac{\partial f}{\partial x_1} &= \frac{\partial f}{\partial x_2} = \frac{\partial f}{\partial x_3} = \bar{x}^2 \\ \frac{\partial^2 f}{\partial x_1^2} &= \frac{\partial^2 f}{\partial x_2^2} = \frac{\partial^2 f}{\partial x_3^2} = 0 \text{ e } \frac{\partial^2 f}{\partial x_1 \partial x_2} = \frac{\partial^2 f}{\partial x_1 \partial x_3} = \frac{\partial^2 f}{\partial x_2 \partial x_3} = \bar{x} \\ \frac{\partial^3 f}{\partial x_1 \partial x_2 \partial x_3} &= 1 \text{ e } \frac{\partial^2 f}{\partial x_i \partial x_j^2} = 0, \quad \forall i \neq j. \end{aligned}$$

Logo, a expansão de tal função em série de Taylor em torno do ponto  $(\bar{x}, \bar{x}, \bar{x}, \bar{y}, \bar{y}, \bar{y})$  fica

$$\begin{aligned} x_1 x_2 x_3 &= \bar{x}^3 + \bar{x}^2(x_1 - \bar{x}) + \bar{x}^2(x_2 - \bar{x}) + \bar{x}^2(x_3 - \bar{x}) \\ &+ \bar{x}(x_1 - \bar{x})(x_2 - \bar{x}) + \bar{x}(x_1 - \bar{x})(x_3 - \bar{x}) + \bar{x}(x_2 - \bar{x})(x_3 - \bar{x}) + \frac{1}{6}(x_1 - \bar{x})(x_2 - \bar{x})(x_3 - \bar{x}) \end{aligned}$$

Agora iremos usar que  $ab = [a^2 + b^2 - (a - b)^2]/2$ , de forma que

$$\begin{aligned} x_1 x_2 x_3 &= \bar{x}^2(x_1 + x_2 + x_3) + \frac{\bar{x}}{2} [(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 - (x_1 - x_2)^2] \\ &+ \frac{\bar{x}}{2} [(x_1 - \bar{x})^2 + (x_3 - \bar{x})^2 - (x_1 - x_3)^2] + \frac{\bar{x}}{2} [(x_2 - \bar{x})^2 + (x_3 - \bar{x})^2 - (x_2 - x_3)^2] \\ &+ \frac{1}{6}(x_1 - \bar{x})(x_2 - \bar{x})(x_3 - \bar{x}) \end{aligned}$$

Logo,

$$\begin{aligned} x_1 x_2 x_3 &= 3\bar{x}^3 + \bar{x} [(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2] \\ &- \frac{\bar{x}}{2} [(x_1 - x_2)^2 + (x_1 - x_3)^2 + (x_2 - x_3)^2] + \frac{1}{6}(x_1 - \bar{x})(x_2 - \bar{x})(x_3 - \bar{x}) \\ &= 3\bar{x}^3 + 3\bar{x}\sigma_x^2 - \frac{3}{2}\bar{x}\bar{\sigma}_x^2 + \frac{1}{6}(x_1 - \bar{x})(x_2 - \bar{x})(x_3 - \bar{x}) \\ &= 3\bar{x}^3 + 3\bar{x}\sigma_x^2 - \frac{9}{2}\bar{x}\sigma_x^2 + \frac{1}{6}(x_1 - \bar{x})(x_2 - \bar{x})(x_3 - \bar{x}) \\ &= 3\bar{x}^3 - \frac{3}{2}\bar{x}\sigma_x^2 + \frac{1}{6}(x_1 - \bar{x})(x_2 - \bar{x})(x_3 - \bar{x}) \end{aligned}$$

Em resumo,

$$\bar{x}_g^3 = 3\bar{x}^3 - \frac{3}{2}\bar{x}\sigma_x^2 + \frac{1}{6}(x_1 - \bar{x})(x_2 - \bar{x})(x_3 - \bar{x})$$

## 7.6 Prova do Lema 2

Por definição

$$\sigma_x^2 = \frac{1}{3} ((x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2)$$

Logo, substituindo o valor de  $\bar{x} = (x_1 + x_2 + x_3)/3$ ,

$$\begin{aligned} \sigma_x^2 &= \frac{1}{27} ((2x_1 - x_2 - x_3)^2 + (2x_2 - x_1 - x_3)^2 + (2x_3 - x_1 - x_2)^2) \\ &= \frac{1}{27} (6x_1^2 + 6x_2^2 + 6x_3^2 - 6x_1x_2 - 6x_1x_3 - 6x_2x_3) \\ &= \frac{6}{27} (x_1^2 + x_2^2 + x_3^2 - (x_1x_2 + x_1x_3 + x_2x_3)) = \frac{2}{9} (x_1^2 + x_2^2 + x_3^2 - 3\hat{x}^2) \end{aligned}$$

Logo,

$$\frac{9}{2}\sigma_x^2 = x_1^2 + x_2^2 + x_3^2 - 3\hat{x}^2 \quad (15)$$

Agora, elevando ao quadrado a média aritmética, temos

$$\bar{x}^2 = \frac{1}{9}(x_1 + x_2 + x_3)^2 = \frac{1}{9}(x_1^2 + x_2^2 + x_3^2 + 2x_1x_2 + 2x_1x_3 + 2x_2x_3)$$

Logo,

$$9\bar{x}^2 = x_1^2 + x_2^2 + x_3^2 + 6\hat{x}^2 \tag{16}$$

Multiplicando a equação em (15) por -1 e somando o resultado com a equação (16), temos

$$-\frac{9}{2}\sigma_x^2 + 9\bar{x}^2 = 9\hat{x} \Leftrightarrow \hat{x}^2 = \bar{x}^2 - \frac{1}{2}\sigma_x^2 \quad \square$$

### 7.7 Prova da Afirmação 1

**Lema 5** Dado  $x = \{x_1, x_2, x_3\}$ , com  $x_i \neq x_j$ , para todo  $i \neq j$ , existe  $y = \{y_1, y_2, y_3\}$ , com  $x_i, y_i \in (0, 1]$ , para todo  $i = 1, 2, 3$ , tais que  $\bar{x} > \bar{y}$ ,  $\sigma_x^2 < \sigma_y^2$  e  $\bar{x}_g < \bar{y}_g$ .

**Prova:** Inicialmente observamos que o lema ainda é verdadeiro se trocarmos  $\sigma_x^2$  por  $\bar{\sigma}_x^2$ , devido a identidade em (1), e faremos isso para facilitar os cálculos. A prova segue da existência de um vetor  $\eta = (a, b, c)$  tal que

$$\begin{cases} \frac{\partial \bar{x}}{\partial \eta} = \langle \text{grad } \bar{x}, \eta \rangle < 0 \\ \frac{\partial \bar{\sigma}_x^2}{\partial \eta} = \langle \text{grad } \bar{\sigma}_x^2, \eta \rangle > 0 \\ \frac{\partial \bar{x}_g}{\partial \eta} = \langle \text{grad } \bar{x}_g, \eta \rangle > 0 \end{cases} \tag{17}$$

Neste caso existiria  $t > 0$ , suficientemente pequeno, tal que  $y = x + t\eta$  satisfaz a tese do lema.

Temos os seguintes gradientes

$$\begin{aligned} \text{grad } \bar{x} &= \frac{1}{3}(1, 1, 1) \\ \text{grad } \bar{\sigma}_x^2 &= 2(2x_1 - x_2 - x_3, 2x_2 - x_1 - x_3, 2x_3 - x_1 - x_2) \\ \text{grad } \bar{x}_g &= \frac{1}{3(x_1x_2x_3)^{2/3}}(x_2x_3, x_1x_3, x_1x_2) \end{aligned}$$

De forma que (17) equivale a

$$\begin{cases} a + b + c < 0 \\ (2x_1 - x_2 - x_3)a + (2x_2 - x_1 - x_3)b + (2x_3 - x_1 - x_2)c > 0 \\ (x_2x_3)a + (x_1x_3)b + (x_1x_2)c > 0 \end{cases} \tag{18}$$

E para mostrar a existência do vetor  $\eta = (a, b, c)$  que satisfaz (18), basta garantir a existência de solução para o sistema linear

$$\begin{cases} a + b + c = -1 \\ (2x_1 - x_2 - x_3)a + (2x_2 - x_1 - x_3)b + (2x_3 - x_1 - x_2)c = 1 \\ x_2x_3a + x_1x_3b + x_1x_2c = 1 \end{cases} \tag{19}$$

que equivale ao determinante abaixo ser não nulo,

$$\det \begin{pmatrix} 1 & 1 & 1 \\ 2x_1 - x_2 - x_3 & 2x_2 - x_1 - x_3 & 2x_3 - x_1 - x_2 \\ x_2x_3 & x_1x_3 & x_1x_2 \end{pmatrix} = 3(x_1 - x_2)(x_2 - x_3)(x_3 - x_1) \neq 0,$$

o que ocorre, pois, por hipótese  $x_i \neq x_j$ , para todo  $i \neq j$ . Lembamos que o determinante ser diferente de zero é condição suficiente para existência de solução do sistema em (19).



O **Texto para Discussão - 139** e outras publicações do IPECE encontram-se disponíveis na internet através do endereço:  
[www.ipece.ce.gov.br](http://www.ipece.ce.gov.br)