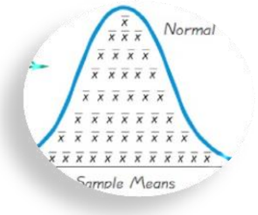
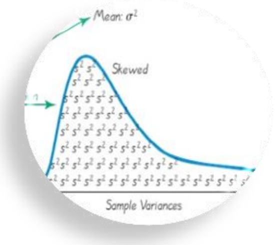


Teoria da estimação



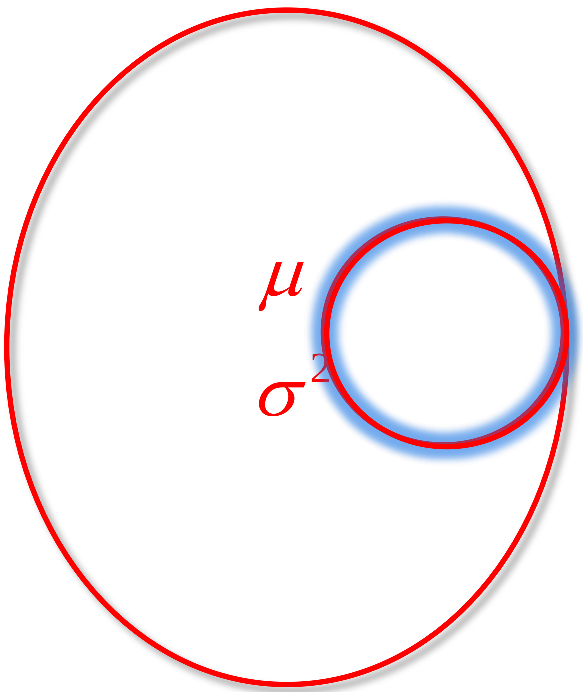
César A. Ticona-Benavente

População (N)

μ

σ^2

População (N)



Amostra (n)

$$A = T(X_1, X_2, \ldots, X_n)$$

Amostragem

$\overline{x}$

s^2

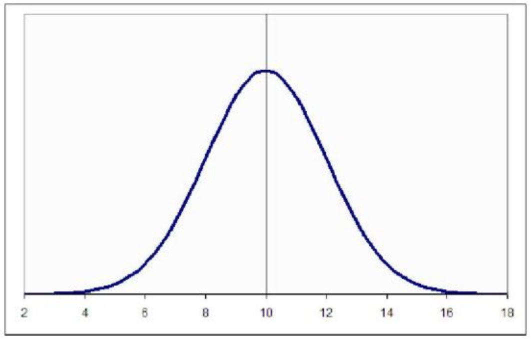
Inferência estatística

Estimadores

$$\overline{x}_1 = f_{x1}(A) = \hat{\mu}_1$$
$$\overline{x}_2 = f_{x2}(A) = \hat{\mu}_2$$
$$\overline{x}_3 = f_{xn}(A) = \hat{\mu}_3$$

Estimativa

Modelo



Problema:

Nunca se sabe o valor de μ e σ^2

Estimação de Parâmetros

População → Amostra

Distribuição de Probabilidade (ou FDP) | Distribuição Amostral

Parâmetros (valor fixo) $\xleftarrow{\text{estimar}}$ Estatísticas (variável aleatória)

Estimação
- pontual (estatísticas)
- por intervalo (intervalos de confiança)

Estatística: é a v.a. que estima (pontualmente) um parâmetro (populacional) as vezes é chamada simplesmente de estimador

Estimativa: é o valor do estimador obtido para uma amostra específica

Estimação

O objetivo é procurar valores que representem adequadamente os parâmetros desconhecidos.

A procura será realizada seguindo algum critério

Para uma distribuição dada só corresponde um único valor do parâmetro.

Como fazer a melhor estimação dos parâmetros?

1. Escolhe o modelo probabilístico adequado

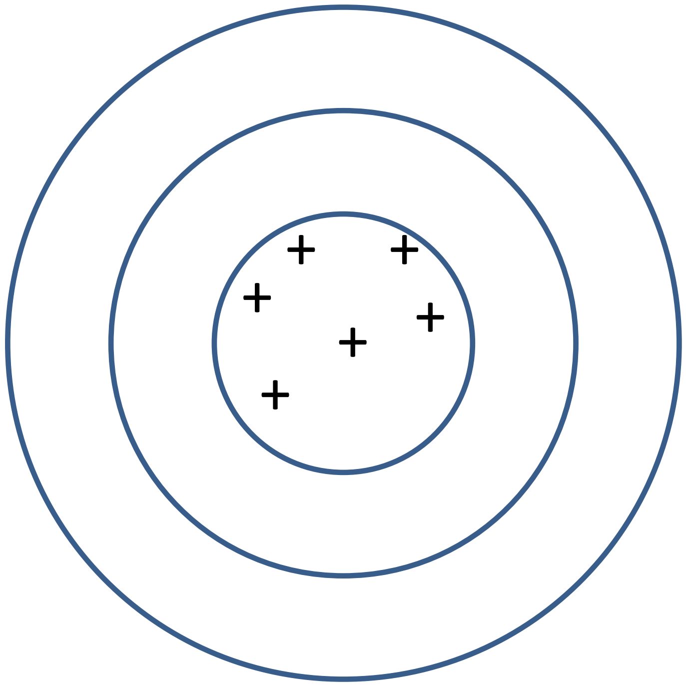
- Normal: Peso, altura, pressão arterial, etc.
- Exponencial: Tempo de vida de equipamentos
- Binomial: Proporção de peças defeituosas
- Poisson: Numero de pessoas que chegam a um local em um determinado intervalo de tempo
- Outros: Uniforme discreto, Gama, etc.

2. Embasado no modelo, escolhe-se o melhor estimador

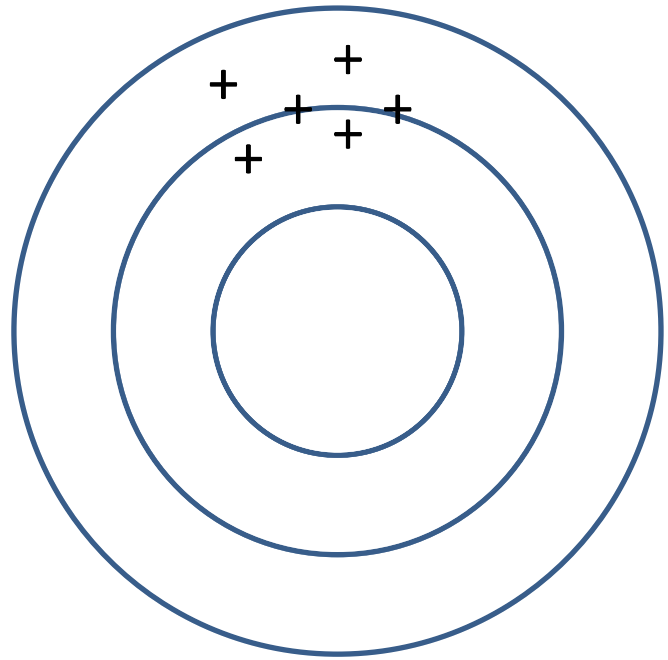
Características dos bons estimadores

- **Não viesado**: Quando a distribuição do estimador esta centrada no parâmetro. $E(\hat{\theta}) = \theta$
- **Eficiente**: É um estimador não viesado que tem a menor variância $Var(\hat{\theta}_1) < Var(\hat{\theta}_2)$
- **Consistente (coerente)**: Quando a distribuição se torna mais concentrada à medida que n tende ao infinito. $i)\ \lim_{n \to \infty} E(\hat{\theta}) = \theta;\quad ii)\ \lim_{n \to \infty} Var(\hat{\theta}) = 0$
- **Suficiente:** Quando uma estatística não utiliza td a amostra, mas é suficiente para estimar o parâmetro.

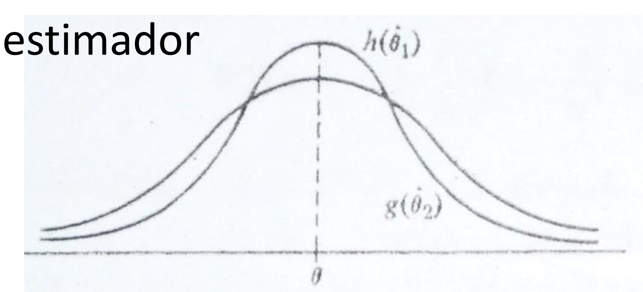
Não viesado

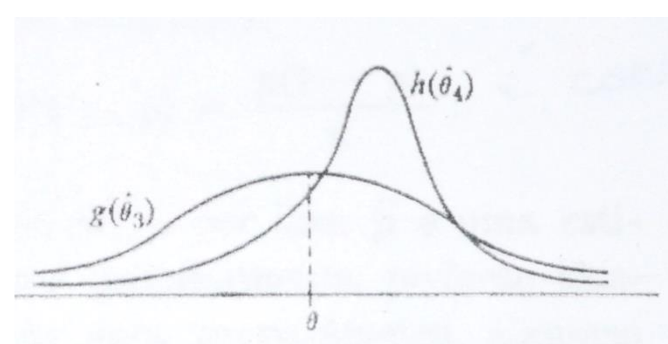


Viesado ou tendencioso

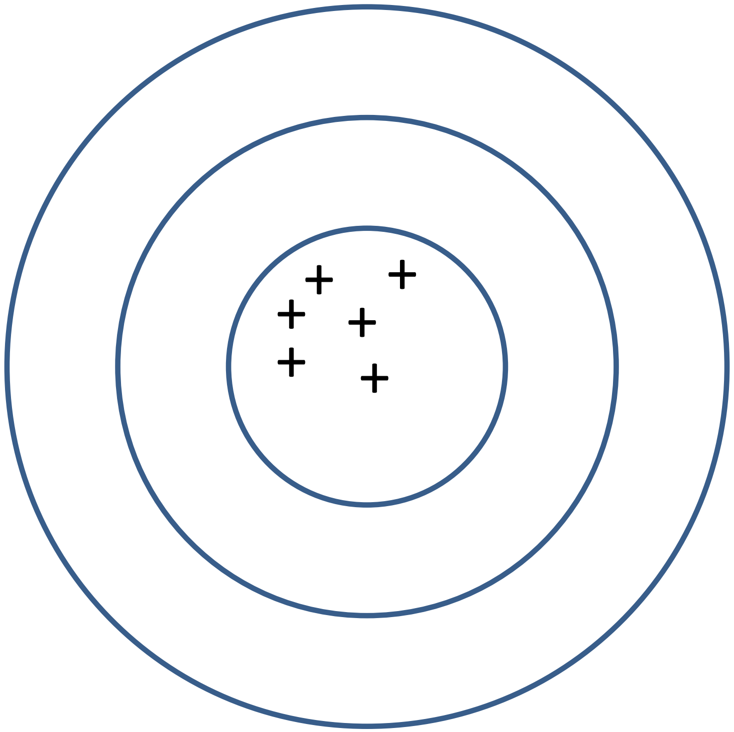


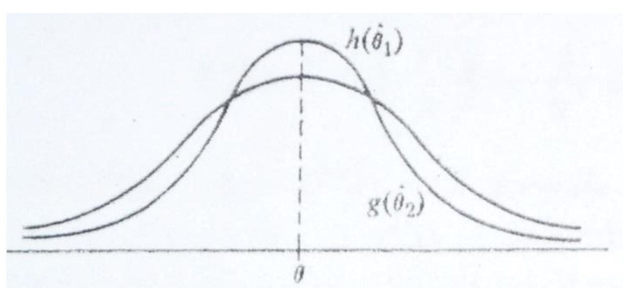
Distrib. do estimador





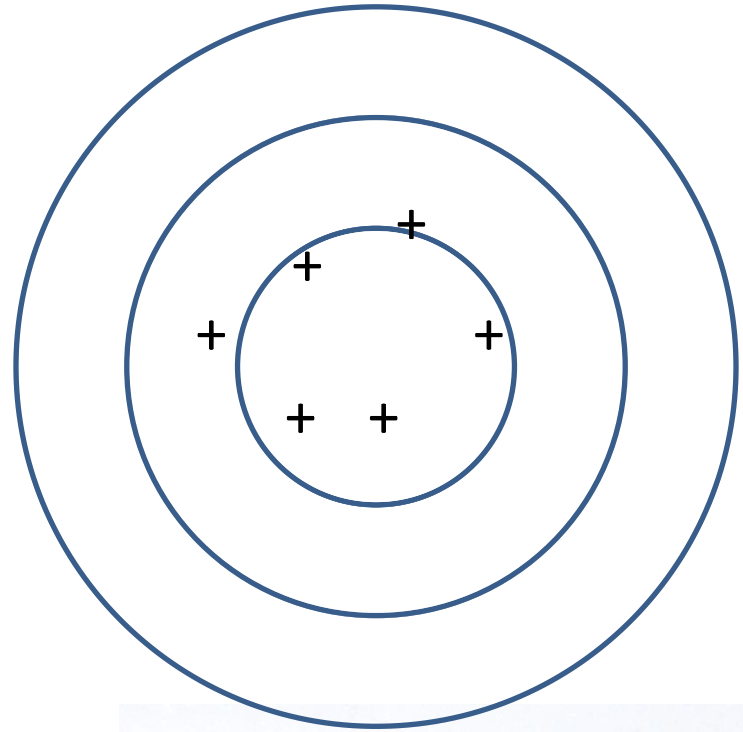
Eficiente

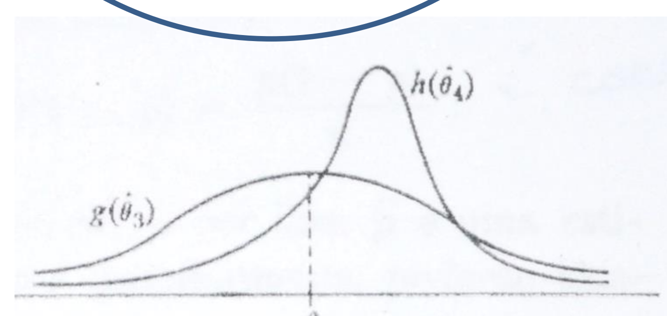




$\hat{\theta}_1$ é mais eficiente

Não Eficiente

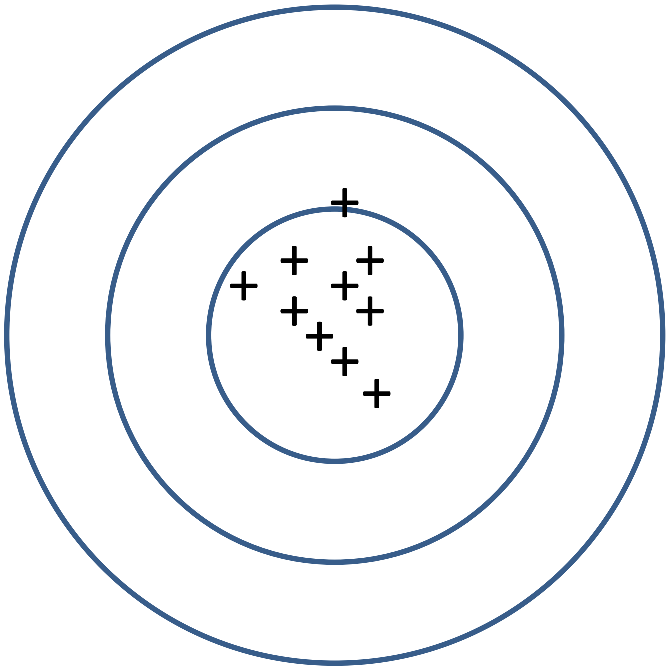




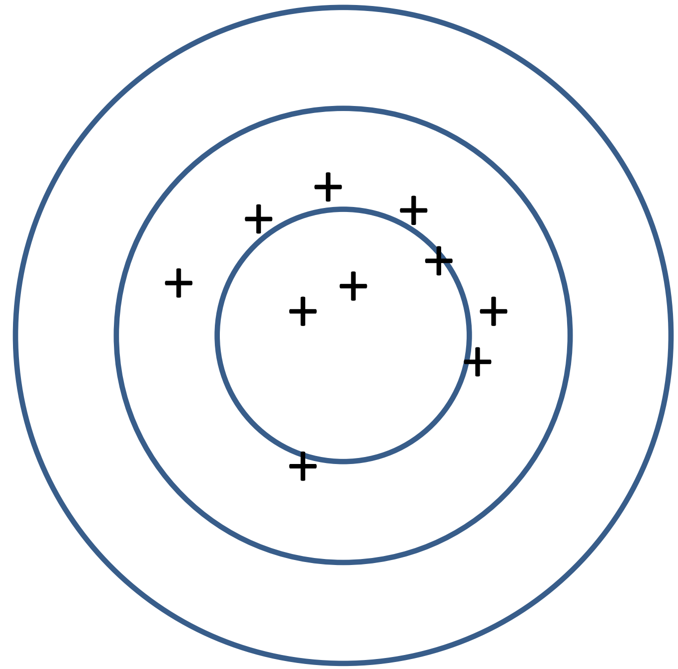
$\hat{\theta}_1$ é viesado, com var minima

$\hat{\theta}_2$ é não é viesado, a var não é min

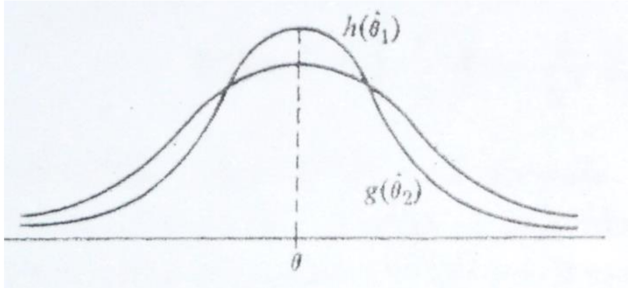
Consistente



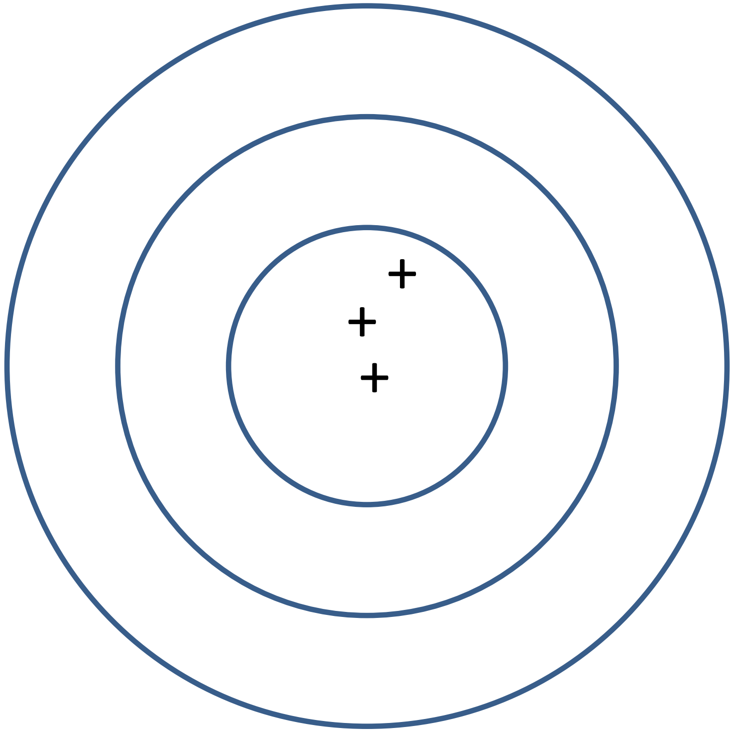
Não Consistente



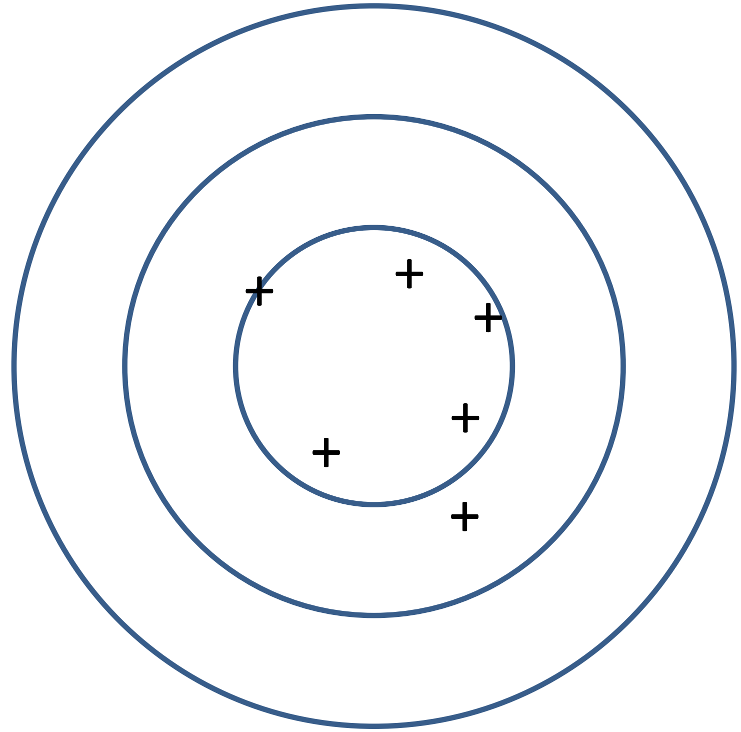
Não é viesado
A var da distrib de $\hat{\theta}_1$
é menor a medida
que n aumenta



Suficiente



Não Suficiente



Exemplo. A população tem 3 elementos (1,2 e 5). São tiradas todas as amostras possíveis de tamanho 2 com reposição

Amostra	media	mediana	amplit	s^2	s	Prop de impares	Prob
1, 1	1.0	1.0	0.0	0.0	0.000	1.0	0.111
1, 2	1.5	1.5	1.0	0.5	0.707	0.5	0.111
1, 5	3.0	3.0	4.0	8.0	2.828	1.0	0.111
2, 1	1.5	1.5	1.0	0.5	0.707	0.5	0.111
2, 2	2.0	2.0	0.0	0.0	0.000	0.0	0.111
2, 5	3.5	3.5	3.0	4.5	2.121	0.5	0.111
5, 1	3.0	3.0	4.0	8.0	2.828	1.0	0.111
5, 2	3.5	3.5	3.0	4.5	2.121	0.5	0.111
5, 5	5.0	5.0	0.0	0.0	0.000	1.0	0.111
Media	2.67	2.67	1.78	2.89	1.26	0.67	
Parâmetro	2.67	2.00	4.00	2.89	1.70	0.67	
Estatística atinge o parâmetro ?	**SIM**	**NÃO**	**NÃO**	**SIM**	**NÃO**	**SIM**	

Viesado vs Não Viesado

Parâmetro	Estimador	Propriedades
μ	$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$	Não viciado e consistente
p	$\hat{p} = \frac{\%\,\text{com característica}}{n}$	Não viciado e consistente
σ^2	$S^2 = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n-1}$	Não viciado e consistente
σ^2	$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n}$	Viciado e consistente

Exemplo

Considere que, numa certa população, uma variável aleatória X assuma os valores 0, 10, 20 e 30 com porcentagens 20%, 30%, 30% e 20%, respectivamente. Logo $\mu = 15$ e $\sigma^2 = 105$.

X	**0**	**10**	**20**	**30**
P(X=x)	**0,2**	**0,3**	**0,3**	**0,2**

Retirando todas as amostras de 2 elementos com reposição tem-se a distribuição conjunta

$X_1 \backslash X_2$	0	10	20	30
0	0,04	0,06	0,06	0,04
10	0,06	0,09	0,06	0,09
20	0,06	0,09	0,09	0,06
30	0,04	0,06	0,06	0,04

Vamos verificar estes estimadores

$$\hat{\mu}_1 = X_1$$

Um numero tomado ao acaso

$$\hat{\mu}_2 = \frac{\sum_{i=1}^{n} X_i}{n}$$

$\hat{\mu}_1$	0	10	20	30
$P(\mu_1=x)$	0,2	0,3	0,3	0,2

$\hat{\mu}_2$	0	5	10	15	20	25	30
$P(\mu_2=x)$	0,04	0,12	0,21	0,26	0,21	0,12	0,04

Por definição $\boxed{E(X) = \Sigma(x.f_x)}$

- Calculando a esperança dos estimadores

$$\mu_1 \begin{cases} E(\hat{\mu}_1) = E\left(X_i\right) = \mu \\ E(\hat{\mu}_1) = 0*0{,}2 + 10*0{,}3 + 20*0{,}3 + 30*0{,}3 = \boxed{15} \end{cases}$$

$$\mu_2 \begin{cases} E(\hat{\mu}_2) = E(X) = E\left(\frac{X_1 + X_2 + \cdots + X_n}{n}\right) = \frac{1}{n}E\left(X_1 + X_2 + \cdots + X_n\right) = \frac{n\mu}{n} = \mu \\ E(\hat{\mu}_2) = 0*0{,}04 + 5*0{,}12 + 10*0{,}21 + \ldots + 30*0{,}04 = \boxed{15} \end{cases}$$

- Por tanto ambos estimadores **não são viciados**

$$E(\hat{\mu}_1) = \mu$$

$$E(\hat{\mu}_2) = \mu$$

- Esses estimadores são consistentes?

Propriedade da variância

$$Var(aX) = a^2 Var(X)$$

$Para -> \hat{\mu}_1$

$$i)\ \lim_{n\to\infty} E(\hat{\mu}_1) = \lim_{n\to\infty} E(X_1) = \lim_{n\to\infty} 15 = \mu$$

$$ii)\ \lim_{n\to\infty} Var(\hat{\mu}_1) = \lim_{n\to\infty} Var(X_1) = \lim_{n\to\infty} 105 = \sigma^2$$

$Para -> \hat{\mu}_2$

$$i)\ \lim_{n\to\infty} E(\hat{\mu}_2) = \lim_{n\to\infty} E\left(\frac{X_1 + X_2 + \ldots + X_n}{n}\right) = \lim_{n\to\infty} \frac{\sum_{i=1}^{n} E(X_i)}{n} = \lim_{n\to\infty} \frac{n\mu}{n} = \mu$$

$$ii)\ \lim_{n\to\infty} Var(\hat{\mu}_2) = \lim_{n\to\infty} Var\left(\frac{X_1 + X_2 + \ldots + X_n}{n}\right) = \lim_{n\to\infty} \frac{\sum_{i=1}^{n} Var(X_i)}{n^2} = \lim_{n\to\infty} \frac{n\sigma^2}{n^2} = \lim_{n\to\infty} \frac{\sigma^2}{n} = 0$$

-O estimador $\hat{\mu}_1$ não é consistente

-O estimador $\hat{\mu}_2$ é consistente

Qual é o estimador mais eficiente?

$$Var(\hat{\mu}_1) = \sigma^2 = 105$$

$$Var(\hat{\mu}_2) = \frac{\sigma^2}{n} = \frac{105}{2} = 52{,}5$$

Portanto o estimador 2 é mais eficiente que o estimador 1.

$$Var(\hat{\mu}_2) < Var(\hat{\mu}_1)$$

Exemplo. Proporção de plantas resistentes

$$Amostra = (X_1, \ldots, X_n)$$ dos quais *k* são resistentes

$$Y = \sum_{i=1}^{n} X_i = k$$

A estimativa do parâmetro *sugestiva é* $\longrightarrow$ $\hat{p} = Y / n$

Verificando se é viesado

$$E(\hat{p}) = E(\tfrac{Y}{n}) = \tfrac{1}{n}(np) = p$$

Isso indica que o estimador de p não é tendencioso

Verificando se é consistente

$$Var(\hat{p}) = Var(\tfrac{Y}{n}) = \frac{Var(Y)}{n} = \tfrac{1}{n^2}(npq) = \tfrac{1}{n^2} np(1-p) = \frac{p(1-p)}{n}$$

Se n tende ao infinito a variância do estimador de p tende a zero.
Portanto, este estimador é consistente

Vejamos outro estimador

A estimativa do parâmetro *sugestiva é* $\longrightarrow$ $\hat{p} = X_i$

Verificando se é viesado

$$E(\hat{p}) = 1P(X = 1) + 0(PX = 0) = p$$

Verificando se é consistente

$$Var(\hat{p}) = pq = p(1 - p)$$

Esta variância é maior que a variância anterior (Não Eficiente)
Quando n tende ao infinito a variância não tende a zero. (Não consistente)

Exemplo. O estimador da variância para distribuição normal é viesado?

Seja X uma v.a. normalmente distribuída com a média (μ) e a variância (σ^2) desconhecidas. Retira-se uma amostra de tamanho n com a finalidade de se estimar μ e σ^2.

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}\left(x_i - \bar{X}\right)^2}{n}$$

$$\boxed{Var(\bar{X}) = \frac{\sigma^2}{n}}$$

$$\boxed{Var(X_i) = E\left(X_i^2\right) - \left(E\left(X_i\right)\right)^2}$$

$$E\left(\hat{\sigma}^2\right) = E\left(\frac{\sum_{i=1}^{n} X_i^2 - n\bar{X}^2}{n}\right) = \frac{1}{n}E\left(\sum_{i=1}^{n} X_i^2\right) - E\left(\bar{X}^2\right) = \frac{1}{n}\sum_{i=1}^{n}\underbrace{E\left(X_i^2\right)}_{\mathrm{I}} - \underbrace{E\left(\bar{X}^2\right)}_{\mathrm{II}}$$

$$= \sigma^2 + \mu^2 - \frac{\sigma^2}{n} - \mu^2 = \frac{n\sigma^2 - \sigma^2}{n} = \frac{n-1}{n}\sigma^2$$

I

$$Var(X_i) = \sigma^2$$

$$\sigma^2 = E\left(X_i^2\right) - \left(E\left(X_i\right)\right)^2$$

$$\sigma^2 = E\left(X_i^2\right) - \mu^2$$

$$E\left(X_i^2\right) = \sigma^2 + \mu^2$$

O esperado era $Var(\hat{\mu}_1) = \dfrac{\sigma^2}{n}$

II

$$\frac{\sigma^2}{n} = E\left(\bar{X}^2\right) - \left(E\left(\bar{X}\right)\right)^2$$

$$\frac{\sigma^2}{n} = E\left(\bar{X}^2\right) - \mu^2$$

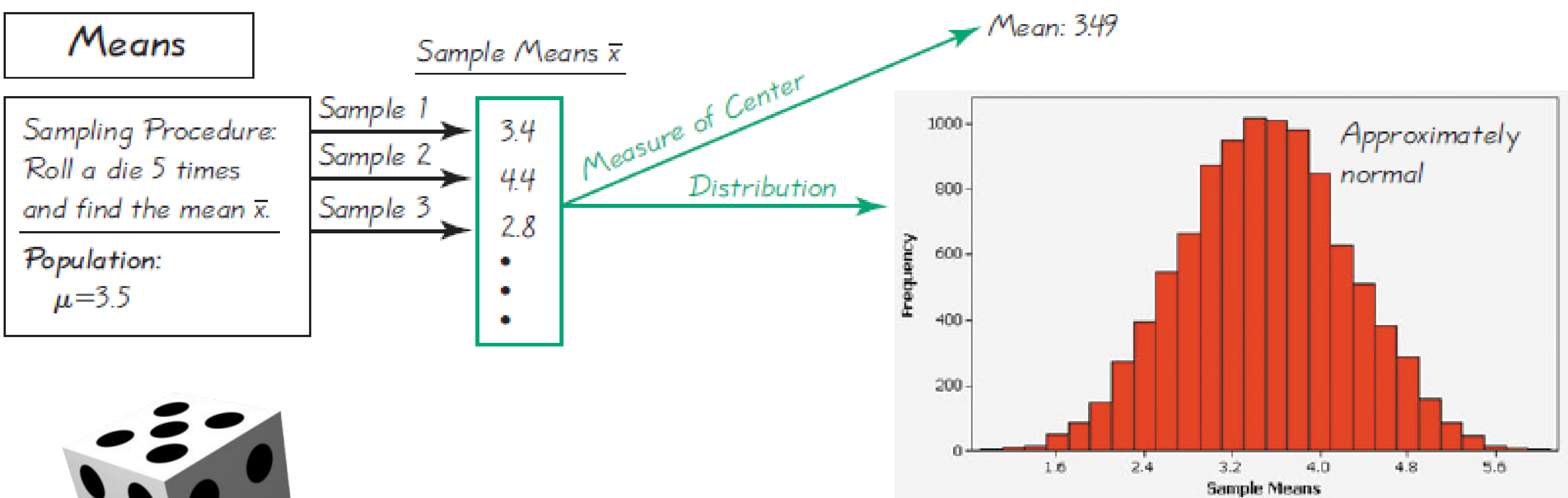
$$E\left(\bar{X}^2\right) = \frac{\sigma^2}{n} + \mu^2$$

Métodos para conseguir os bons estimadores

- Método da máxima verossimilhança
- Método dos momentos
- Método dos quadrados mínimos

PARA PENSAR

- A média do dado é 3,5.
- É retirada 10000 amostras de tamanho 5 com reposição.
- A média das 10000 médias amostrais é 3.49

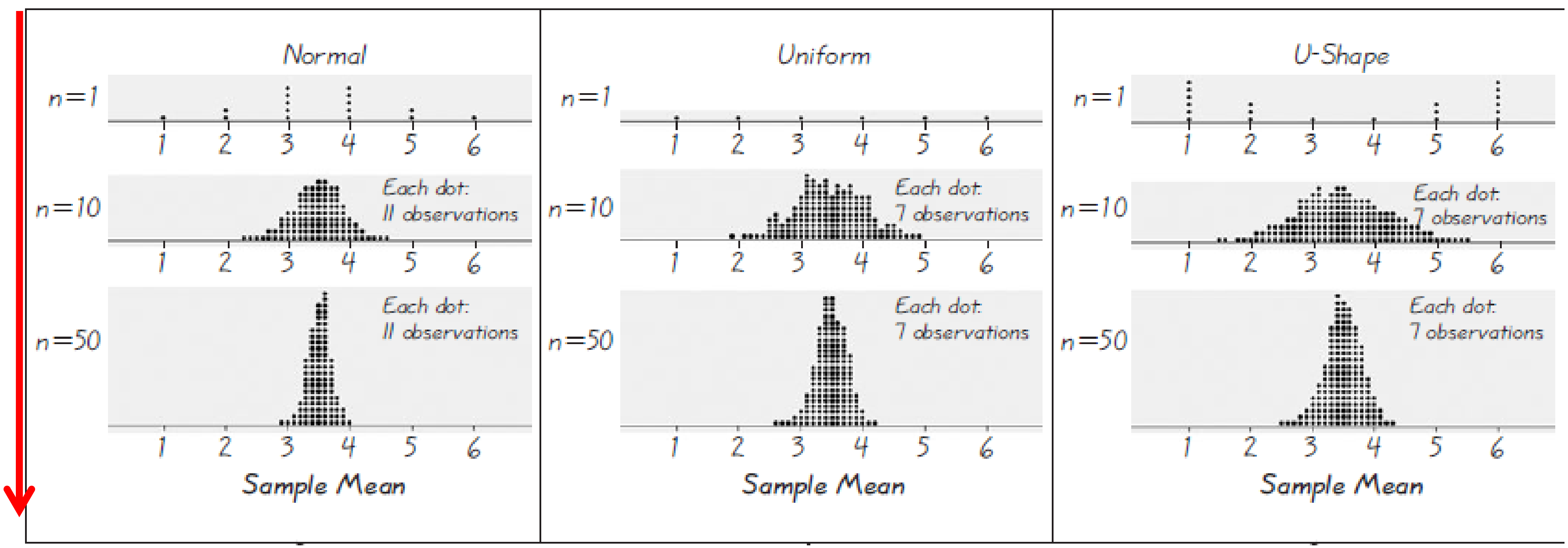


A DISTRIBUIÇÃO DAS MEDIAS
AMOSTRAIS É NORMAL

Teorema central do limite

Distribuição X → Amostras maiores → Distribuição NORMAL

Table 6-6 Sampling Distributions



A distribuição das médias amostrais é normal

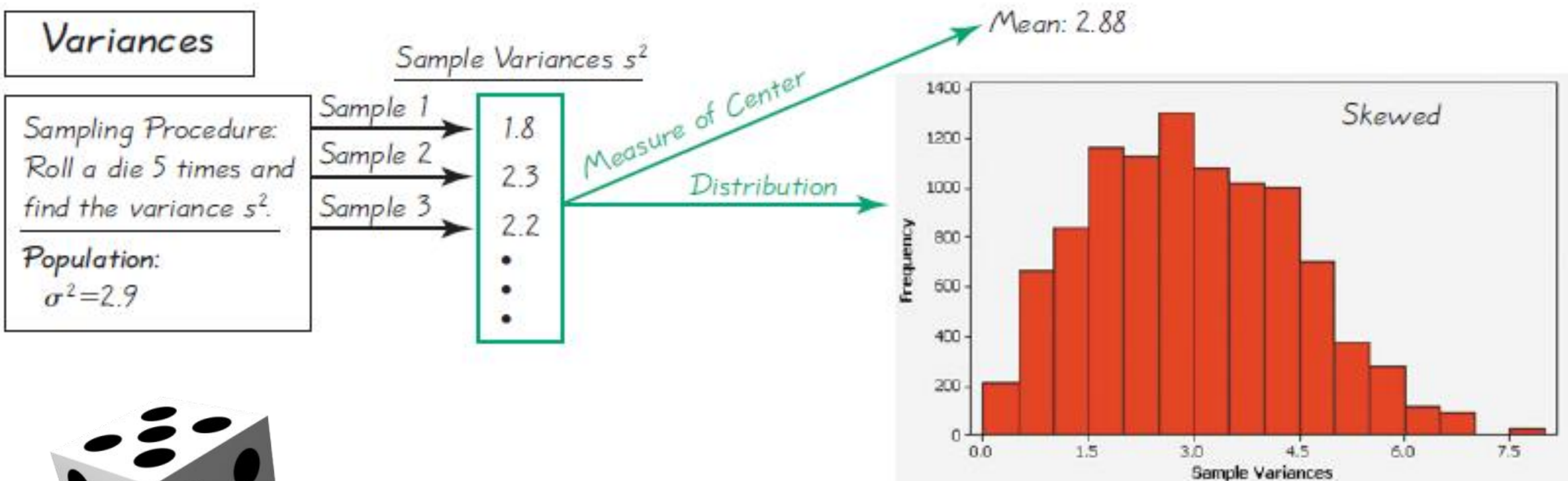
Conclusões interessantes à partir do teorema

- Se a população original não for normal, para n>30 a distribuição das médias amostrais pode ser aproximada pela distr. NORMAL.
- A aproximação é melhor à medida que o tamanho de amostra aumenta
- Se a população for normalmente distribuída, então as médias amostrais serão normalmente distribuídas para qualquer tamanho da amostra.

VARIANCIA

PARA PENSAR

- A variância populacional do dado é 2,9.
- São retiradas 10000 amostras de tamanho 5 com reposição.
- A média das 10000 variâncias amostrais é 2.88

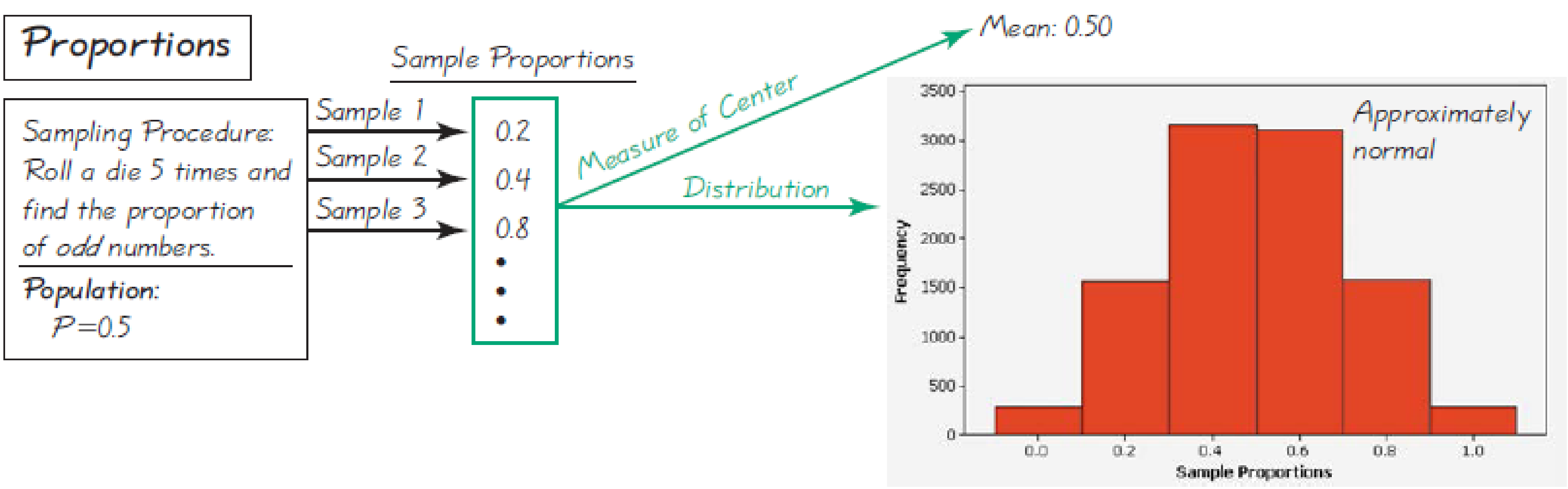


A DISTRIBUIÇÃO DAS VARIANCIAS AMOSTRAIS É ASSIMETRICA (X^2)

PROPORÇÕES

PARA PENSAR

- A proporção de números impares no dado é 0.5
- São retiradas 10000 amostras de tamanho 5 com reposição.
- A média das 10000 proporções amostrais é 0.5



A DISTRIBUIÇÃO DAS PROPORÇÕES AMOSTRAIS É NORMAL

Também pode-se verificar

$$\sigma_{\overline{x}} = \frac{\sigma}{\sqrt{n}}$$

n=tamanho amostral

$$\sigma_{\overline{x}} = \frac{\sigma}{\sqrt{n}} = \frac{1.70}{\sqrt{2}} = 1.2019$$

$$\sigma_{\overline{x}} = DP_{popul}(1, 1.5, 3, 1.5, \ldots, 5) = 1.2019$$

Este valor também é chamado de
erro padrão da média