Estimação

Prof. Adriano Mendonça Souza, Dr.
Departamento de Estatística
PPGEMQ / PPGEP - UFSM

Estimação de Parâmetros

O objetivo da Estatística Indutiva é tirar conclusões probabilísticas sobre aspectos da população, observando amostras dessa população.

Podemos subdividir a Estatística Indutiva em dois grandes grupos: os problemas de estimação e os testes de hipótese.

Estimação

É o processo que usa os resultados extraídos da amostra para testar valores de certos parâmetros da população (testes paramétricos) ou para testar a natureza da distribuição da população (testes não-paramétricos).

- Se $\underline{\theta}$ for um parâmetro conhecido, podemos calcular todas as probabilidades através do modelo $p(x/\underline{\theta})$.

- Como não conhecemos $\underline{\theta}$ precisamos mensurar alguns elementos da população, digamos $\underline{x}=(x_1,x_2,\ldots,x_n)$ e,

- A partir desses valores realizar inferências, sobre $\underline{\theta}$ e outras características populacionais $(\mu,\sigma^2,\ldots)$

- O vetor $\underline{\theta}=(\theta_1,\theta_2,\ldots,\theta_p)$ é composto por quantidades desconhecidas não aleatórias

- Considere $\underline{x} = (x_1, x_2, \ldots, x_n)$ n VA's, que formam uma amostra, com função densidade de probabilidade conjunta $p(x_1, x_2, \ldots, x_n / \underline{\theta}) = p(\underline{x} / \underline{\theta})$

- A partir dessa amostra, e sua fdp conjunta, é que podemos "ESTIMAR" os parâmetros desconhecidos $\underline{\theta}$ (estatisticamente)

Para isso necessitamos de:

- Teoria das distribuições amostrais

  D. Normal; D. Qui-quadrado; D. “t” de Student; D. “F” de Snedecor

- Métodos de Estimação

  M. dos Momentos; M. dos Mín. quadrados; M. Máx. Verossimilhança e propriedades estatísticas dos estimadores

- Teorias de testes de Hipóteses

Parâmetros e Estatísticas

Qualquer valor calculado com base nos elementos de uma amostra é chamado uma "estatística"

Parâmetro	Estimador
μ = média σ^2 = variância σ = desvio padrão	$\bar{x}$ = média s^2 = variância s = desvio padrão

Inferência Estatística

Uma população pode ser inferida através de sua amostra através de dois modos:

- **Estimação**
- **Testes de hipóteses**

Estimação

A estimação de um parâmetro populacional pode ser de dois tipos:

estimação pontual

estimação intervalar

Conceitos de Estimador e Estimativa

Estimador ($\widehat{\theta}$) de um parâmetro populacional teta é uma variável aleatória função dos elementos amostrais $\widehat{\theta} = f(X_1, X_2, \ldots, X_n)$;

Estimativa é o valor numérico obtido pelo estimador (ou estatística em uma determinada amostra.

Principais qualidades de um estimador

Quando propomos $\hat{\theta}$ como uma estimativa de θ, não esperamos realmente que $\hat{\theta}$ venha a ser igual a θ. (Recordemos que é uma VA. e, por isso, pode tomar diferentes valores). Este dilema origina duas importantes questões:

1. Quais as características que desejamos que uma "boa" estimativa apresente?

2. Como decidiremos que uma estimativa é "melhor" do que a outra?

Tais questões não admitem uma análise tão simples.

1) o próprio valor do parâmetro θ não é verdadeiramente correto (ao menos em uma situação concreta). Assim, como dizer que $\hat{\theta}$ o é?

2) se tivermos duas estimativas de $\hat{\theta}$,

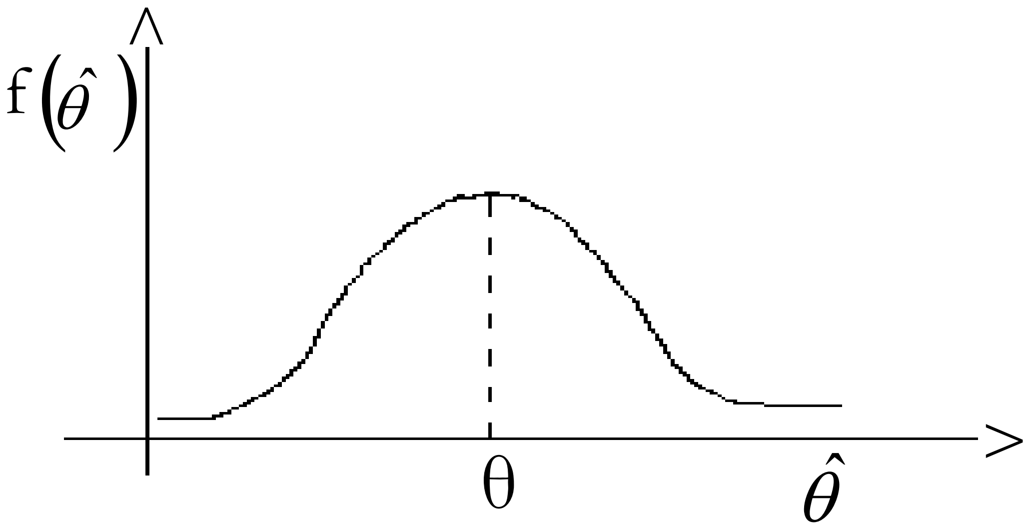
por exemplo $\hat{\theta}_0$ e $\hat{\theta}_1$, qual o critério de comparação para decidirmos entre $\hat{\theta}_0$ e $\hat{\theta}_1$?

Justeza (ou não-tendenciosidade, ou sem vício,ou sem viés)

Nós diremos que $\hat{\theta}$ é uma estimativa não tendenciosa de θ se:

$$\boxed{E(\hat{\theta}) = \theta \quad \forall \theta}$$

Estimador Não-tendencioso:



$E[\hat{\theta}] = \theta \Rightarrow$ A média da V.A.

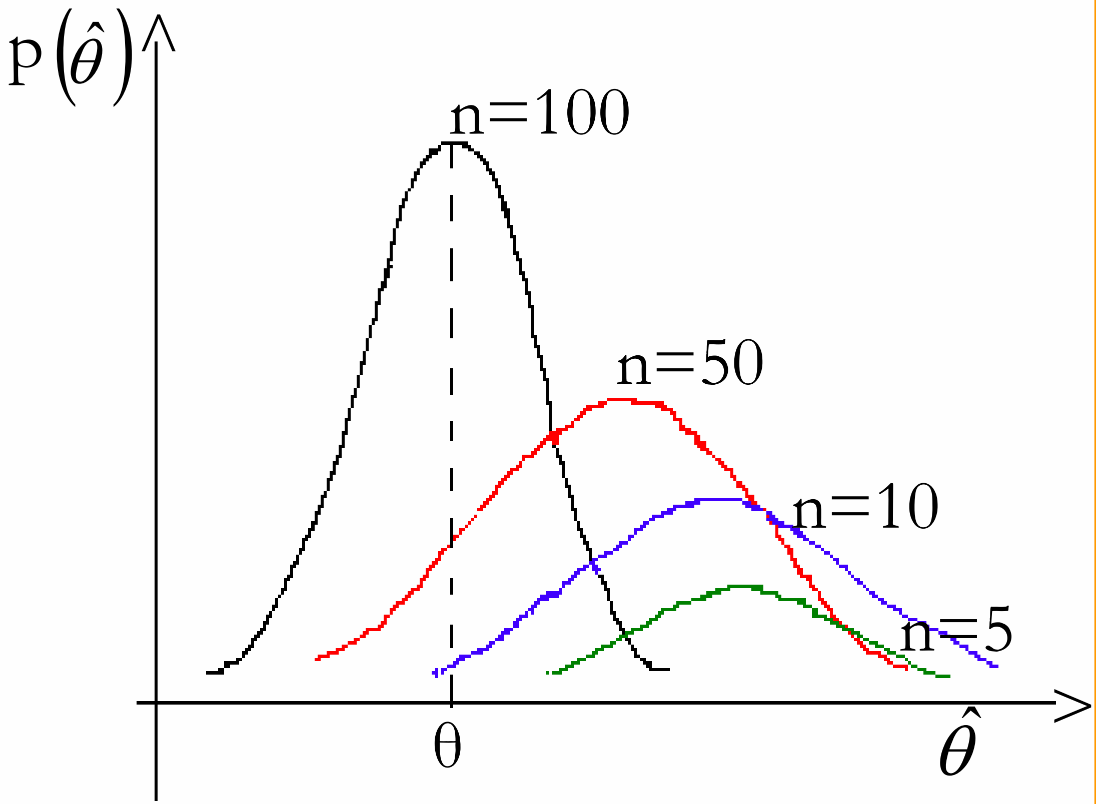
$\hat{\theta}$ coincide com θ ,o verdadeiro valor do parâmetro.

Consistência (ou coerência)

$\hat{\theta}$ é consistente se (além de ser justo) sua variância tende para zero, quando *n* é suficientemente grande (n $\rightarrow\propto$), isto é:

$$\boxed{E\left[\hat{\theta}\right]=\theta \qquad e \qquad Lim_{\,n\rightarrow\infty}\, Var\left[\hat{\theta}\right]=0}$$

Estimador Consistente:



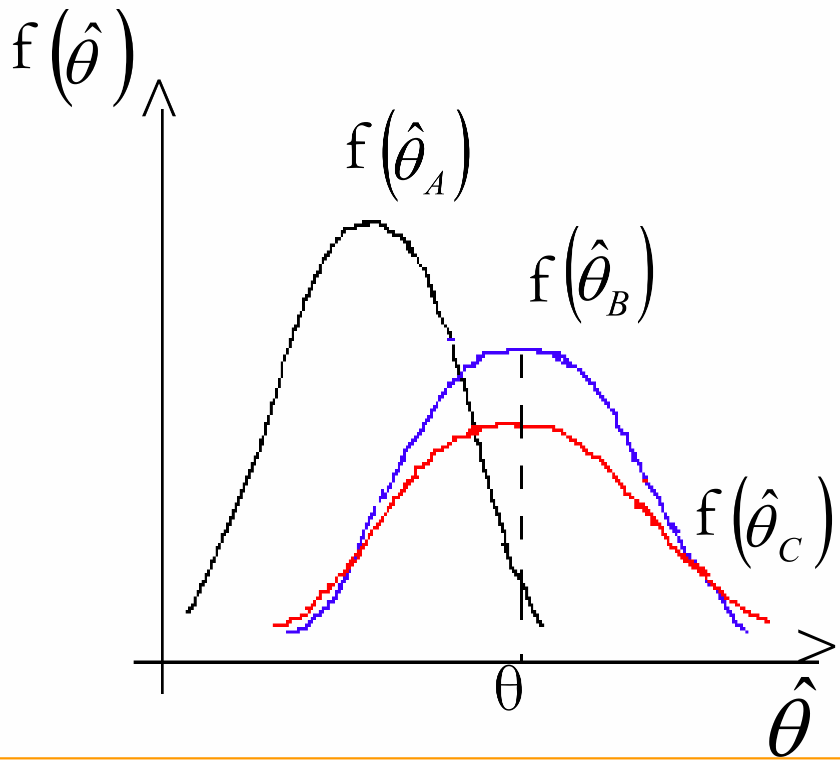
* **A propriedade da consistência garante que nossas estimativas melhoram à medida que o tamanho da amostra aumente.**

3) Eficiência

Seja $\hat{\theta}$ uma estimativa não-tendenciosa de θ.

Diremos que é uma estimativa de variância mínima de θ, se para todas as estimativas $\hat{\theta}^*$ tais que $E(\hat{\theta}^*) = \theta$, tivermos $Var(\hat{\theta}) \leq Var(\hat{\theta}^*)$ para todo θ .

Estimador Eficiente:



* $\hat{\theta}_A, \hat{\theta}_B, \hat{\theta}_C$ são estimadores de θ.

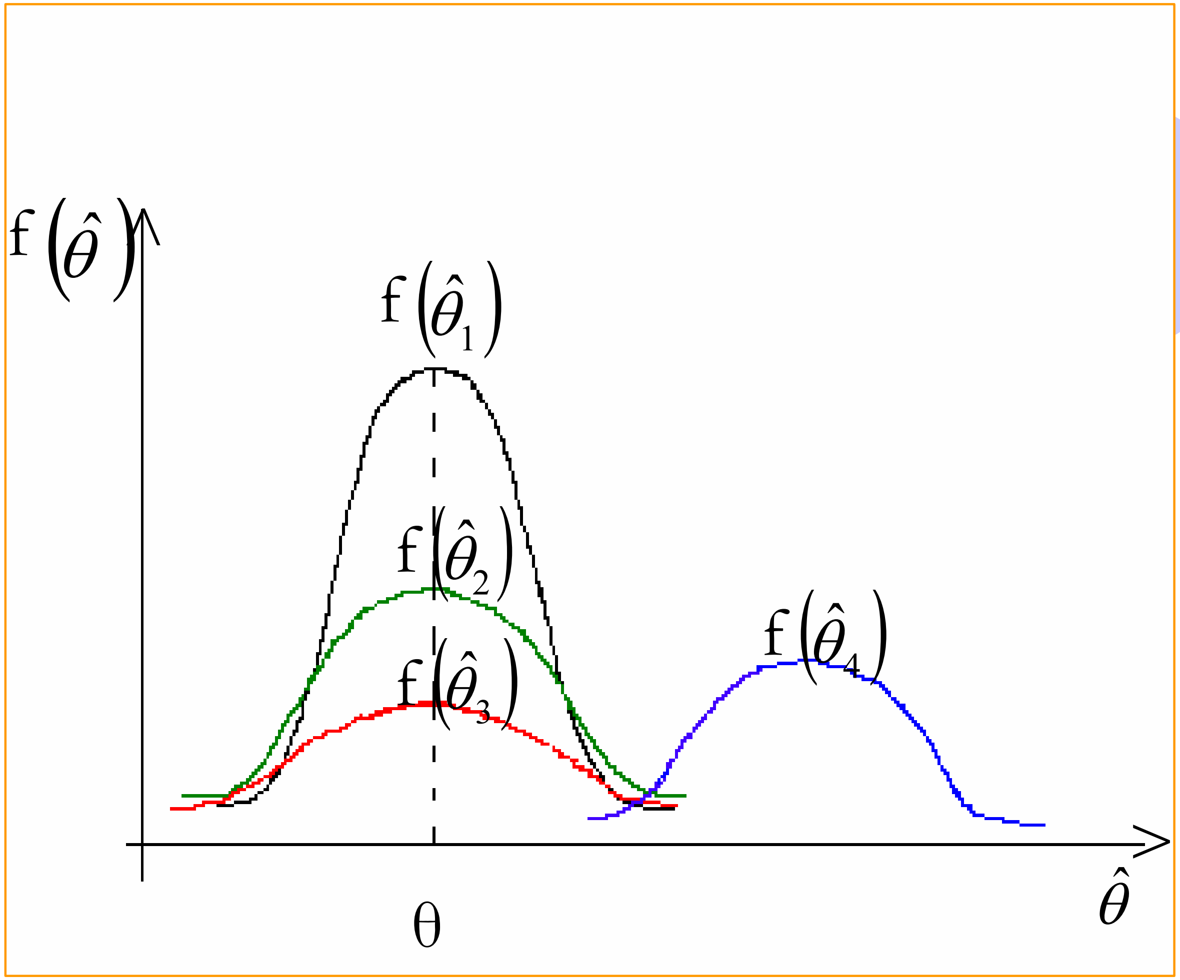
* $\hat{\theta}_B, \hat{\theta}_C$ são não tendenciosos.

* $\hat{\theta}_A$ é tendencioso, mas com variância mínima.

* $\hat{\theta}_B$ será um estimador eficiente de $\hat{\theta}$ se não houver nenhum outro estimador não tendencioso de $\hat{\theta}$ com menor variância.

Exemplo:

Observando o gráfico abaixo, responda:



a) Quais os estimadores justos de θ?

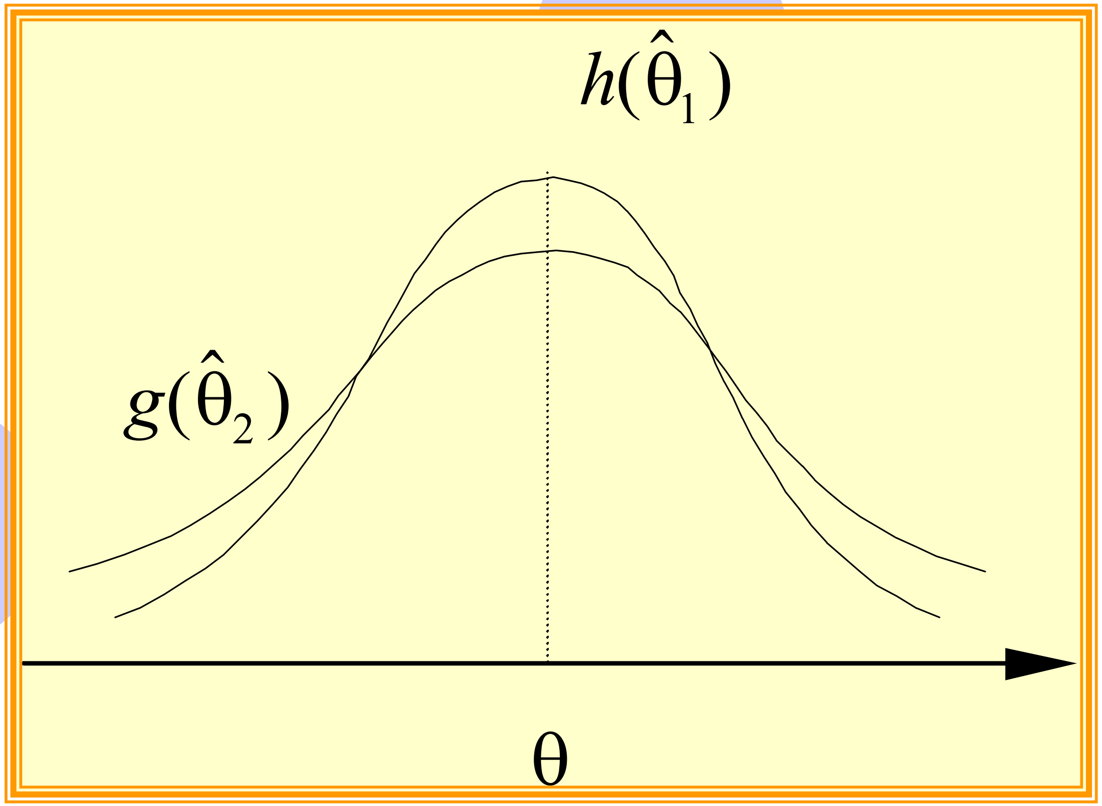
b) Quais os estimadores viezados de θ?

c) Qual o estimador de variância mínima de θ?

d) Qual o estimador de maior variância?

e) Qual deles você escolheria?

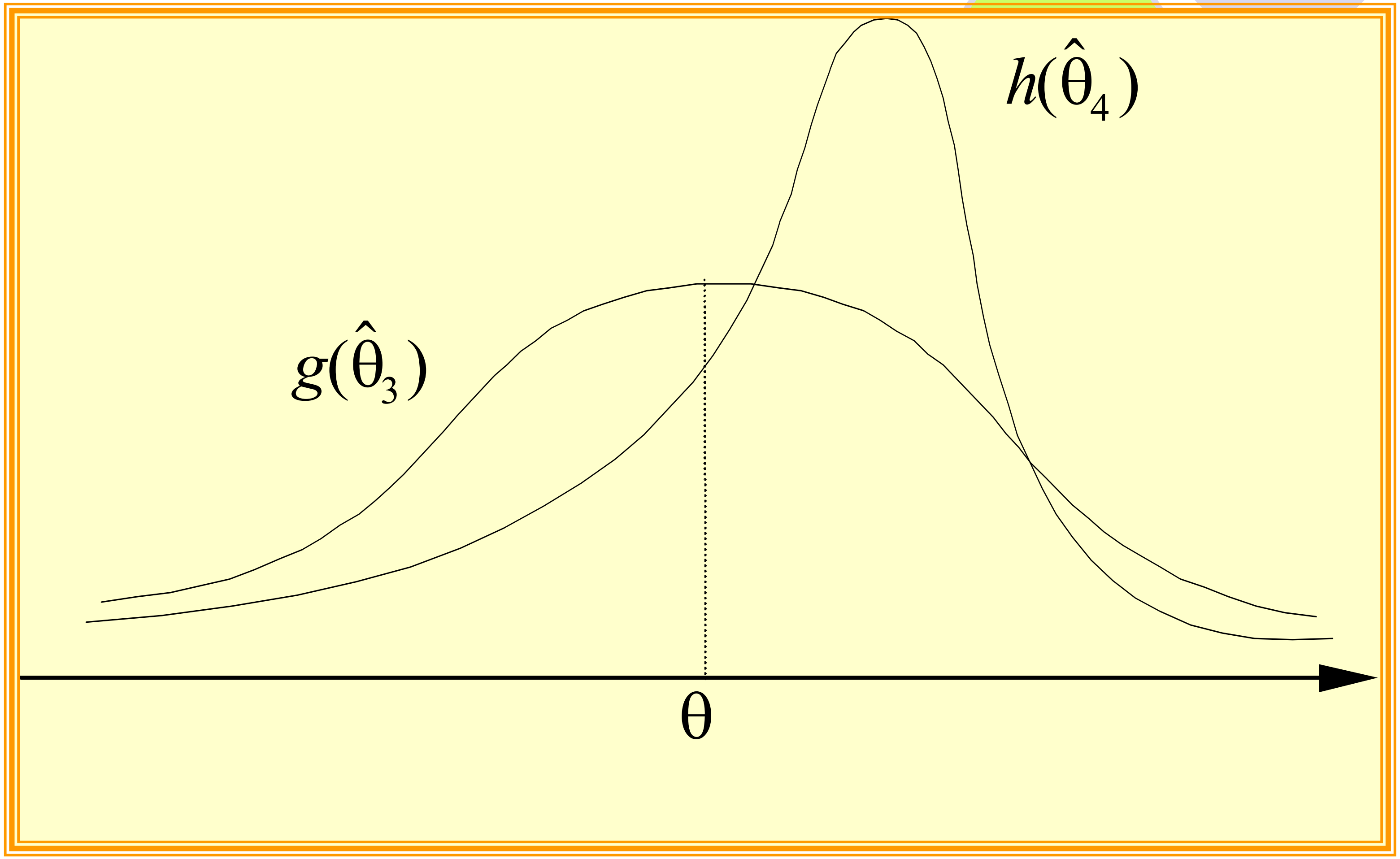
Assim, se $\hat{\theta}_1$ e $\hat{\theta}_2$ forem duas estimativas de θ, cuja *fdp* seja representada na figura seguinte, presumivelmente preferiríamos $\hat{\theta}_1$ a $\hat{\theta}_2$.



No caso de estimativas $\hat{\theta}_3$ e $\hat{\theta}_4$, (ver figura a seguir) a decisão não é tão evidente, porque $\hat{\theta}_3$ é não-tendenciosa, enquanto $\hat{\theta}_4$ é tendenciosa.

Todavia $Var(\hat{\theta}_3) > Var(\hat{\theta}_4)$. Isto significa que, enquanto em média $\hat{\theta}_3$ será próxima de θ, sua variância revela que desvios consideráveis em relação a θ não serão de surpreender. $\hat{\theta}_4$ por sua vez, tende a ser um tanto maior do que θ, em média, e no entanto poderá ser mais próxima de θ do que $\hat{\theta}_3$.

Estimativas $\hat{\theta}_3$ e $\hat{\theta}_4$:



Estimação de Parâmetros:

Por pontos: É quando utilizamos um único valor para caracterizar uma população.

Ex.:

$$\overline{X} \rightarrow \mu; \quad S \rightarrow \sigma^2; \quad S^2 \rightarrow \sigma^2$$

Por intervalos: Neste caso utilizamos um intervalo $\hat{\theta}_1 < \theta < \hat{\theta}_2$ com uma certa probabilidade de conter o verdadeiro parâmetro populacional θ .

$$P(\hat{\theta}_1 \leq \theta \leq \hat{\theta}_2) = 1 - \alpha$$

Principais Casos de Intervalos de Confiança:

- Média;
- Proporção;
- Variância;
- Desvio Padrão;
- Soma e diferença de médias;
- Soma e diferença de proporções;
- Quociente das variâncias.