Introdução à Estatística

1. Probabilidade

Espaço Amostral (Ω) : Enumeração (finita ou infinita) de todos os resultados possíveis.

$\Omega = A1, A2, A3, \ldots$

Evento (A): Resultados ou conjunto de resultados possíveis. Chamamos 'evento' qualquer subconjunto do espaço amostral.

Evento Impossível (ø): Conjunto Vazio, pois ele nunca acontecerá.

Probabilidade (P(A)): Probabilidade de um evento **A** ocorrer.

$$P(A) = \frac{A}{\Omega}$$

União - (A ∪ B)

> Pelo menos um ocorre

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

$P(A \cup B) = P(A) + P(B)$, Para eventos mutuamente exclusivos.

Interseção - (A ∩ B)

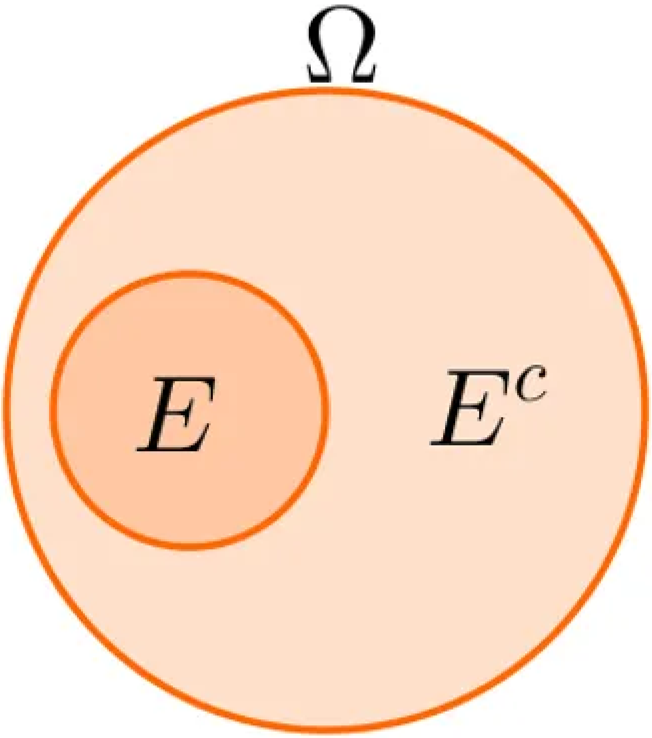
> A *e* B ocorrem>

$P(A \cap B) = P(A) \cdot P(B)$ - Eventos Independentes

$P(A \cap B) = P(A) \cdot P(B|A)$ - Eventos Dependentes

$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$

Evento Complementar (A^c)



$P(A^c) = 1 - P(A)$

$(A \cap B)^c = A^c \cup B^c$

$(A \cup B)^c = A^c \cap B^c$

Probabilidade Condicional (B|A)

> B dado que A ocorre

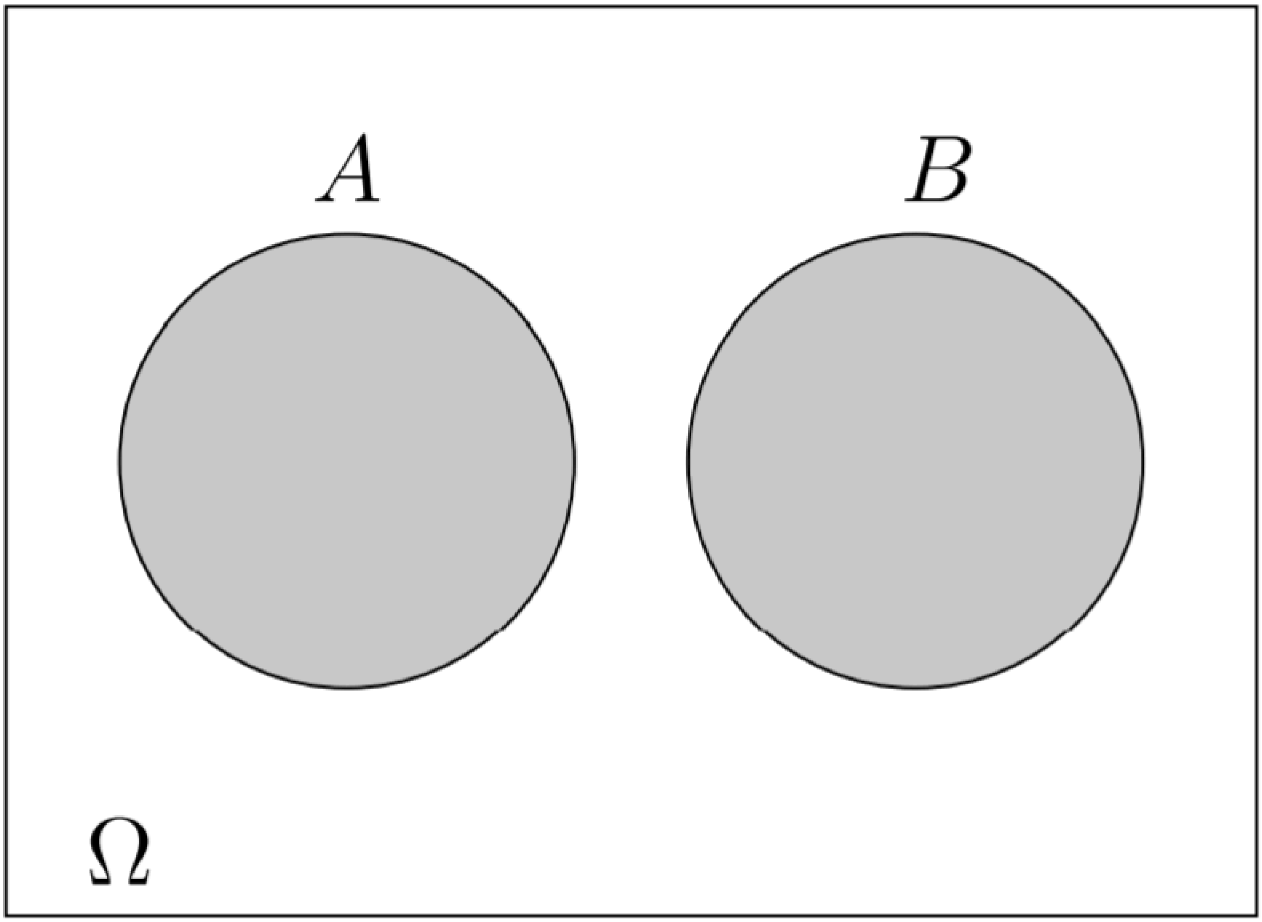
$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

$P(A|B) = 1 - P(A^c|B)$

$P(A_1 \cup A_2|B) = P(A_1|B) + P(A_2|B) - P(A_1 \cap A_2|B)$

Eventos Independentes

> A não interfere em B



$P(A|B) = P(A)$

$P(A \cap B) = P(A) + P(B)$

Lei de Morgan

$A^c \cap B^c = (A \cup B)^c = 1 - P(A \cup B)$

$A^c \cup B^c = (A \cap B)^c = 1 - P(A \cap B)$

Teorema de Bayes

$P(A|B) = \frac{(P(A) \cdot P(B|A))}{P(B)}$

$P(A \cap B) = P(A) \cdot (B|A)$

2. Variáveis Aleatórias

> Uma variável aleatória (v.a.) pode ser entendida como uma variável quantitativa, cujo o resultado (valor) depende de fatores aleatórios

$$X : w \in \Omega \to X(w) \in \mathbb{R}$$

$$\text{Ex: } X = \begin{cases} 0, & \text{se ocorrer } \{(C,C)\} \\ 1, & \text{se ocorrer } \{(C,K),(K,C)\} \\ 2, & \text{se ocorrer} \{(K,K)\} \end{cases}$$

Campo de Definição (R_x) = Conjunto de valores possíveis da variável aleatória X, $Rx = (0, 1, 2)$

Variáveis Aleatórias Discretas

> Uma v.a. é discreta se os possíveis resultados estão contidos em um conjunto finito ou enumerável

Função de Probabilidade: associa a cada valor possível da variável aleatória discreta suas respectiva probabilidade

$$p(x) = P(X = x)$$

Tal que,

$$p(x) = \begin{cases} P(X = x), & x \in Rx \\ 0, & \text{caso contrário} \end{cases}$$

Satisfazendo, $p(x) \geq 0$ e $\sum_{x \in Rx} p(x) = 1$

Função de Distribuição Acumulada (FDA)

$$F(x) = P(X \leq x)$$

Onde $F(x)$ é a probabilidade da v.a. X assumir um valor menor ou igual ($\leq$) a x

Satisfazendo, $F(x)$ não é decrescente, $\lim_{x \to -\infty} F(X) = 0$, $\lim_{x \to +\infty} F(X) = 1$

- **Valor Esperado (Esperança)**

$$[E] = \sum_{x \in Rx} x \cdot p(x)$$

Onde, x é o valor de X, e $p(x)$ é a probabilidade de X.

O valor esperado é uma **constante**

- **Variância**

$$Var(X) = \sum_{x \in Rx} (x - E[X])^2 \cdot p(x) = \sum_{x \in Rx} x^2 \cdot p(x) - (E[X])^2$$

Ou seja,

$$Var(X) = E[X^2] - (E[X])^2$$

- Desvio Padrão

$$DP(X) = \sqrt{Var(X)}$$

Modelo Probabilísticos

Modelo Uniforme Discreto

$$X \sim Uniforme\{X_1, X_2, X_3, \ldots, X_n\}$$

$$p(x) = \begin{cases} \frac{1}{x}, & x \in Rx \\ 0, & \text{Caso contrário} \end{cases}$$

- Valor Esperado

$$E[X] = \frac{1}{x} \centerdot \sum_{i=1}^{n} X_i$$

- Variância

$$Var(X) = \frac{1}{x}\left(\sum x^2 - \frac{\left(\sum x\right)^2}{k}\right)$$

Modelo Bernoulli

> Sucesso ou Fracasso

$$X \sim Bernoulli(p) \qquad\qquad \text{, p}(0 < p < 1)$$

$$P(0) = P(X = 0) = 1 - P$$
$$P(1) = P(X = 1) = P$$

$E[X] = P \qquad Var(X) = P(1 = P)$

Modelo Binomial

$X \sim Binomial(n, p)$
Chama-se de experimento binomial ao experimento que

- consiste em n ensaios de Bernoulli
- cujo ensaios são independentes, e
- para qual a probabilidade de sucessos em casa ensaio é sempre igual a p $(0 < p < 1)$

$$P(X = x) = \binom{n}{k} \centerdot p^x \centerdot (1 - p)^{n-x}, \quad k \in \{0, 1, 2, \ldots, n\}$$
$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

$E[X] = n \centerdot p$

$Var(X) = n \centerdot p \centerdot (1 - p)$

Modelo Geométrico

$X \sim Geo(p)$

> Número de repetições de um ensaio de Bernoulli com probabilidade de sucesso p $(0 < p < 1)$até ocorrer o primeiro sucesso

$$p(x) = \begin{cases} p \cdot (1-p)^x - 1, & x \in \mathbb{N} \\ 0, & \text{caso contrário} \end{cases}$$

$E[X] = \frac{1}{p}$
$Var(X) = \frac{1-P}{p^2}$

Modelo Hipergeométrico

$X \sim Hipergeométrico(N, r, n)$

> N $\rightarrow$ Tamanho total
>
> r $\rightarrow$ Número de Casos com atributos de interesse
>
> n $\rightarrow$ Tamanho da amostra

$$p(X) = \begin{cases} \frac{\binom{r}{x} \cdot \binom{N-r}{n-x}}{\binom{N}{n}}, & x \in R_x \\ 0, & \text{Caso contrário} \end{cases}$$

$E[X] = n \cdot \frac{r}{N}$

$Var(X) = n \cdot \frac{n}{N} \cdot \frac{N-r}{N} \cdot \frac{N-n}{N-1}$

Modelo Poisson

$X \sim Poiss(\lambda)$

> Eventos Raros

$$p(x) = e^{-\lambda} \cdot \frac{\lambda^x}{x!}, \quad x = \{0, 1, 2, ..\}$$

$E[X] = Var(x) = \lambda$

Variáveis Aleatórias Contínuas

> Uma $f(x)$ definida sobre o espaço amostral (Ω) e assumindo valores num intervalo de número reais, é dita uma variável aleatória contínua. Uma v.a. é contínua se existir $F_x : \mathbb{R} \rightarrow \mathbb{R}$, denominada **função de densidade de probabilidade (f.d.p.)**, satisfazendo:

$$f(x) \geq 0, \forall x \in \mathbb{R}$$

$$P(a \leq x \leq b) = \int_a^b f(x)dx = P(A) \int_{-\infty}^{+\infty} f(x)dx = 1$$

> Variável Aleatória Discreta $\rightarrow$ **Contagem**
> Variável Aleatória Contínua $\rightarrow$ Medição

- Função de Distribuição Acumulada (f.d.a.)

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt, \quad x \subset \mathbb{R} \ \forall f(x) = F'(x)$$

$E[X] = \int_{-\infty}^{+\infty} x \cdot f(x)dx$

$Var(X) = \int_{\mathbb{R}} (x - E[X])^2 \cdot f(x)dx = E[X^2] \cdot E[X]^2$

Modelos Probabilísticos

Modelo Uniforme Contínuo

$X \sim Uniforme(a, b)$

> Dizemos que X é uma variável uniforme no intervalo $[a.b], (a, b) \in \mathbb{R}, a < b$, se a função de densidade de probabilidade da variável x é constante nesse intervalo e nula fora dele

- Função de densidade de probabilidade

$$f(x) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b \\ 0, & \text{caso contrário} \end{cases}$$

- Função de distribuição

$$F(x) = \begin{cases} 0, & x \leq a \\ \frac{x-a}{b-a}, & a \leq x \leq b \\ 1, & x \geq b \end{cases}$$

- Esperança

$$E[X] = \frac{a+b}{2}$$

- Variância

$$Var(x) = \frac{(a+b)^2}{12}$$

Modelo Exponencial

$X \sim Exp(\lambda)$

> Dizemos que X é uma variável exponencial com parâmetro λ, $\lambda > 0$, se a função de densidade de X é dada por:

$$f(x) = \begin{cases} \lambda \cdot e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

- Função de distribuição

$$F(x) = \begin{cases} 0, & x < 0 \\ 1 - e^{-\lambda x}, & x \geq 0 \end{cases}$$

- Esperança

$$E[X] = \frac{1}{\lambda}$$

- Variância

$$Var(X) = \frac{1}{\lambda^2}$$

Modelo Normal

$X \sim N(\mu, \sigma^2)$

$$f(x) = \frac{1}{\sigma \cdot \sqrt{2\pi}} \cdot e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})}, \quad \forall x \in \mathbb{R}$$

- Esperança

$$E[X] = \mu$$

- Variância

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

3. Variáveis aleatórias bivariadas e n-variadas. Covariância e correlação

Exemplo 1: Considere um teste do tipo certo ou errado com apenas três questões. Suponha que as respostas às questões desse teste serão escolhidas ao acaso. Defina as variáveis

X_1 : número de acertos nas duas primeiras questões, e

X_2 : número de acertos nas duaś ultimas.

Observe que X_1 e X_2 têm o mesmo campo de definição, a saber, $\{0, 1, 2\}$.

Nesse caso, podemos construir uma tabela de dupla entrada....

$X_1 \downarrow X_2 \rightarrow$	0	1	2
0	1/8	1/8	0
1	1/8	1/4	1/8
2	0	1/8	1/8

Probabilidade Marginal

Dada a função de probabilidade conjunta das variáveis X_1 e X_2, como determinar as probabilidades marginais (individuais) das duas variáveis separadamente, isto é, como determinar. por exemplo, $P(X_1 = 2)$?

Nesse caso, tem-se:

$$(X_1 = 2) = (X_1 = 2; X_2 = 0) \cup (X_1 = 2; X_2 = 1) \cup (X_1 = 2; X_2 = 2)$$

ou seja,

$$P(X_1 = 2) = \sum_{x_2 \in R_2} P(X_1 = 2; X_2 = x_2) = \frac{2}{8} = \frac{1}{4}$$

Portanto as funções de probabilidade marginais de X_1 e de X_2 são dadas, respectivamente, por

$$\begin{array}{ll} pX_1(x_1) = \sum_{x_2 \in R_2} p(x_1, x_2), & x_1 \in \mathbb{R} \\ pX_2(x_2) = \sum_{x_1 \in R_2} p(x_1, x_2), & x_2 \in \mathbb{R} \end{array}$$

No exemplo 1, as funções de probabilidade marginais são dadas por:

$X_1 \downarrow X_2 \rightarrow$	0	1	2	px_1
0	1/8	1/8	0	1/4
1	1/8	1/4	1/8	1/2
2	0	1/8	1/8	1/4
px_2	1/4	1/2	1/4	1

$pX_1(x) = pX_2(x)$, qualquer que seja $x \in R$.

Função de probabilidade condicional

Se a função de probabilidade conjunta de X_1 e X_2´e dada por $p(x_1, x_2)$ e os respectivos campos de definição são R_1 e R_2, então a função de probabilidade condicional de X_2 dado que $X_1 = x_1$, $x_1 \in R_1$ é dada por

$$p_{X_2|X_1=x_1}(x_2|x_1) = \frac{P(X_1 = x_1; X_2 = x_2)}{P(X_1 = x_1)}$$

da mesma forma, tem-se

$$p_{X_1|X_2=x_2}(x_1|x_2) = \frac{P(X_1 = x_1; X_2 = x_2)}{P(X_2 = x_2)}$$

Exemplo 2: Determine a probabilidade condicional de X_2 dado que $X_1 = 0$

$X_1 \downarrow X_2 \rightarrow$	**0**	**1**	**2**	px_1
0	1/8	1/8	0	1/4
1	1/8	1/4	1/8	1/2
2	0	1/8	1/8	1/4
px_2	1/4	1/2	1/4	1

$$P(X_2 = 0; X_1 = 0) = \frac{1/8}{1/4} = \frac{1}{2}$$

$$P(X_2 = 1; X_1 = 0) = \frac{1/8}{1/4} = \frac{1}{2}$$

$$P(X_2 = 2; X_1 = 0) = \frac{0}{1/4} = 0$$

Assim, a função de probabilidade condicional de X_2 dado X_1 = 0 é dada por

x_2	$p_{X_2\|\|X_1=0}$
0	1/2
1	1/2

Independência

Dizemos que X_1 e X_2 são variáveis aleatórias independentes se, e somente se, sua função de probabilidade conjunta fatora no produto de suas funções de probabilidade marginais, isto é.

$$p(x_1, x_2) = p_{X_1}(x_1) \cdot p_{X_2}(x_2), \quad \forall (x_1, x_2) \in \mathbb{R}^2$$

Funções de Variáveis Aleatórias

Se X_1 e X_2 são variáveis aleatórias com função de probabilidade conjunta dada por $p(x_1, x_2)$, podemos definir novas variáveis aleatórias tais como $Z = X_1 + X_2$, $W = X_1 \cdot X_2$ etc.

Exemplo: Defina a variável $W = X_1 \cdot X_2$. Determine o valor esperado de W.

$X_1 \downarrow X_2 \rightarrow$	**0**	**1**	**2**	px_1
0	1/8	1/8	0	1/4

$X_1 \downarrow X_2 \rightarrow$	0	1	2	px_1
1	1/8	1/4	1/8	1/2
2	0	1/8	1/8	1/4
px_2	1/4	1/2	1/4	1

O campo de definição de W e {0, 1, 2, 4} e, as respectivas probabilidades são

(X_1, X_2)	$p(X_1; X_2)$	$W = X_1 \centerdot X_2$
(0,0)	1/8	0
(0,1)	1/8	0
(0,2) = (2,0)	0	0
(1,0)	1/8	1
(1,1)	1/4	1
(1,2)	1/8	2
(2,1)	1/8	2
(2,2)	1/8	4

$W = X_1 \centerdot X_2$	0	1	2	4
$p(W = X_1 \centerdot X_2)$	3/8	2/8	2/8	1/8

$$E[W] = \frac{5}{4} \neq E[X_1 \centerdot X_2] = E[X_1] \centerdot E[X_2]$$

Se definirmos $Z = X_1 + X_2$, o campo de definição de Z é $0, 1, 2, 3, 4$

Assim, diferentemente de $E[W]$:

$$E[Z] = 2 = E[X_1 + X_2] = E[X_1] + E[X_2]$$

Covariância e Correlação

Sejam X_1 e X_2 variáveis aleatórias com função de probabilidade conjunta dada por $p(x_1, x_2)$.
A **covariância** entre X_1 e X_2 é definida por

$$Cov(X_1, X_2) = E[(X_1 - E[E_1]) \centerdot (X_2 - E[X_2])]$$

$$Cov(X_1, X_2) = E[X_1 X_2] - E[X_1]E[X_2]$$

$$E[X_1 X_2] = \sum_{x_1} \centerdot \sum_{X_2} \centerdot p(X_1; X_2)$$

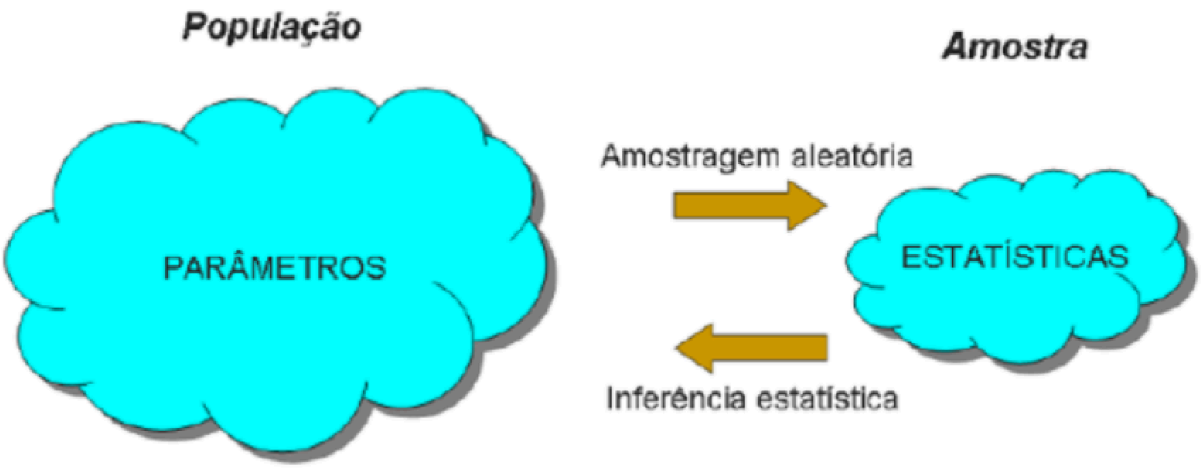
Se as variáveis são independentes, então $Cov(X_1, X_2) = 0 \neq$ Se a $Cov(X_1, X_2) = 0$, as variáveis são independentes.

Sejam X_1 e X_2 variáveis aleatórias com função de probabilidade conjunta dada por $p(x_1, x_2)$.
A **correlação** entre X_1 e X_2 é definida por

$$\rho = \rho_{12} = \frac{Cov(X_1; X_2)}{\sqrt{Var(X_1) \,.\, Var(X_2)}}$$

4. Inferência Estatística

- **População** é o conjunto de todos os elementos sob investigação com pelo menos uma característica em comum.
- **Amostra** é qualquer subconjunto não-vazio da população.



- **Parâmetro** Característica numérica da população
- **Estatística** Característica numérica da população

Atenção: Na estatística inferencial, a palavra estatística tem outro significado.
Um estimador de um parâmetro é uma estatística.

Notação usual para parâmetros e estatísticas

Parâmetro	Estatística
tamanho da população N	tamanho da amostra n
média populacional $\mathrm{E}[X] = \mu$	média amostral $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$
variância populacional $\mathrm{Var}(X) = \sigma^2$	variância amostral $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$
proporção populacional π ou p	proporção amostral $\hat{\pi}$ ou $\hat{p}$

Problemas de Inferência

- Verificação de um tempo médio de vida de uma lâmpada fluorescente especificado pelo fabricante
- Avaliação de um novo produto, Antes do lançamento o produto será distribuído a um grupo de consumidores potenciais que responderão um questionário.
- Previsão do tempo médio de espera dos clientes no caixa do banco.
- Há razões para supor que o tempo de reação Y a certo estímulo visual depende da idade do indivíduo.

Como selecionar uma amostra?

As observações contidas numa amostra são tanto mais informativas sobre a população, quanto mais conhecimento tivermos dessa mesma população.

Por exemplo a análise quantitativa de glóbulos brancos obtida de algumas gotas de sangue da ponta do dedo de um paciente dá a ideia geral da quantidade de glóbulos brancos no corpo todo, pois sabe-se que a distribuição dos glóbulos brancos é homogênea, e de qualquer lugar que se tivesse retirado a amostra ela seria “representativa”.

Nem sempre a escolha de uma amostra adequada é imediata.

Procedimentos de levantamento de dados

- Levantamentos Amostrais

A amostra é obtida de uma população bem definida, por meio de processos bem protocolados e controlados pelo pesquisador.

Tais levantamentos costumam ser subdivididos em dois subgrupos: **probabilísticos e não-probabilísticos**.

O primeiro reúne todas as técnicas que usam mecanismos aleatórios de seleção dos elementos de uma amostra, atribuindo a cada um deles, uma probabilidade, conhecida a priori, de pertencer à mostra.

No segundo grupo estão os demais procedimentos, tais como amostras intencionais, nas quais os elementos s˜ao selecionados com o auxílio de especialistas, e amostras de voluntários, como corre em alguns testes sobre novos medicamentos e vacinas.

- Planejamento de Experimentos

Têm como principal objetivo analisar o efeito de uma variável sobre outra(s). Requer interferências do pesquisador sobre o ambiente em estudo (população), bem como o controle de fatores externos, com o intuito de medir o efeito desejado.

- Levantamentos Observacionais

Os dados são coletados sem que o pesquisador tenha controle sobre as informações obtidas, exceto eventualmente sobre possíveis erros grosseiros. As séries de dados temporais são exemplos típicos desses levantamentos.

Amostragem Aleatória Simples (AAS)

Uma amostra aleatória simples ocorre quando atribuímos probabilidades de seleção na amostra iguais para todos os elementos da população. Com relação a precisão neste tipo de amostragem existe diferença se a seleção é feita com reposição ou sem reposição.

Amostragem Simétrica

Supõe-se dispor de uma listagem de todos os elementos da população em alguma ordem que não esteja relacionada à variável de interesse. Por exemplo, ordem alfabética, ordem de número de matrícula etc.

Suponha que a população tenha N elementos e que iremos sortear uma amostra sistemática de tamanho n, usando essa listagem em que todos os elementos da população est˜ao ordenados de 1 até N.

A ideia é primeiro dividir a listagem em n blocos de tamanhos $k = [N/n]$ em que $[N/n]$ é o menor inteiro que é maior ou igual a N/n.

Amostragem aleatória estratificada

A população é dividida em estratos (subpopulações), geralmente de acordo com os valores (ou categorias) de uma variável, e depois AAS e utilizada na seleção de uma amostra de cada estrato. O procedimento mais comum envolve, depois de fixado o tamanho da amostra, especificar os tamanhos amostrais em cada estrato de forma proporcional ao tamanho de cada estrato.

Amostragem por conglomerados

A população é dividida em grupos (subpopulações) distintos, chamados conglomerados. Por exemplo, podemos dividir uma cidade em bairros ou quadras ou ruas. Usamos AAS para selecionar uma amostra desses conglomerados e depois todos os indivíduos dos conglomerados selecionados são investigados.

Distribuição Amostral

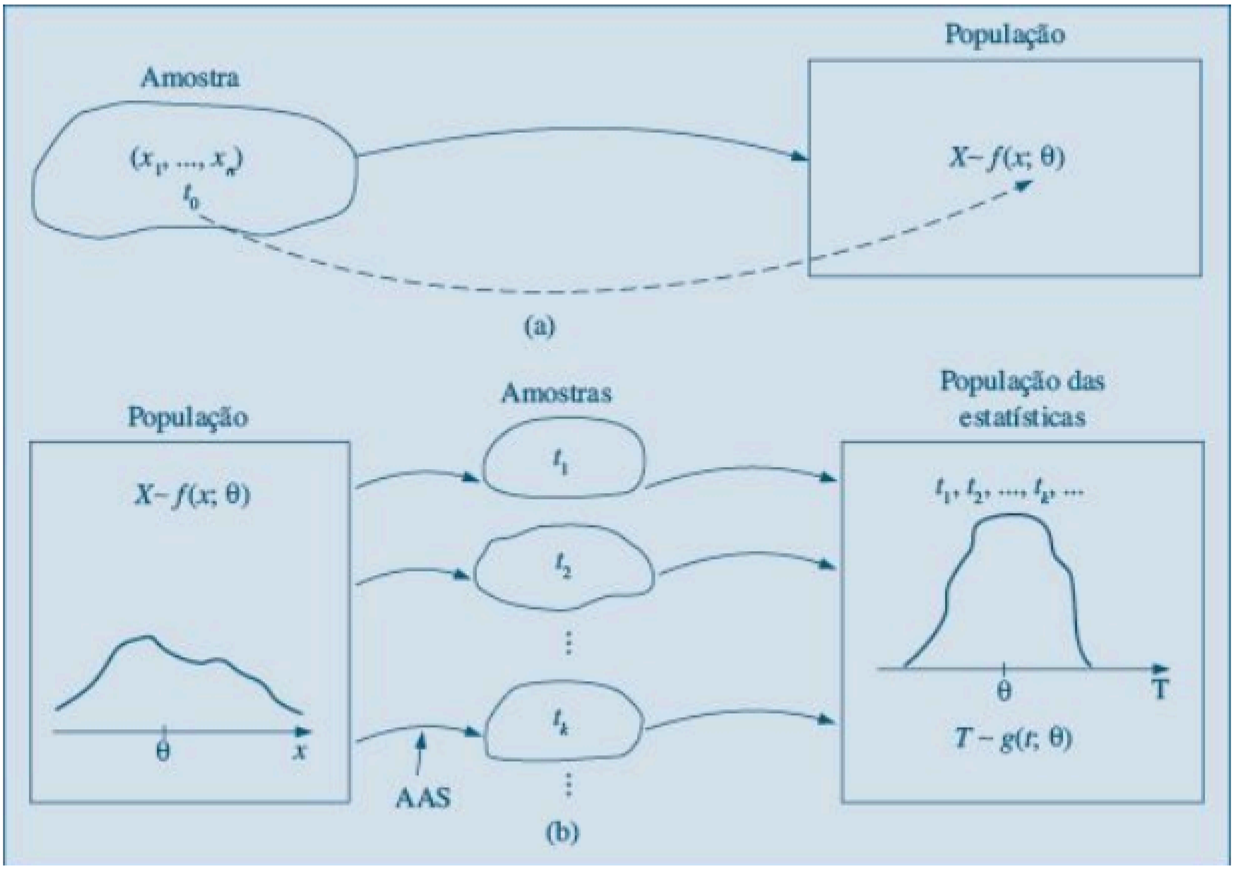
Suponha o problema de estimar um parâmetro θ de certa população e que para isso dispomos de uma amostra de tamanho n dessa população: $x_1, x_2, \ldots, x_n$. Suponha também que usaremos uma estatística T função da amostra para estimar θ.

$$T = t(x_1, x_2, \ldots, x_n)$$

T pode ser a soma $(\sum_{i=1}^{n} x_i)$, a média $(\overline{x})$, a mediana, a amplitude, o desvio padrão amostral, e sua escolha dependerá do parâmetro que queremos estimar.

Essa distribuição é chamada **distribuição amostral da estatística** T e desempenha papel fundamental na teoria da inferência estatística. Esquematicamente, teríamos o procedimento representado abaixo, em que temos:

1. uma população X, com determinado parâmetro de interesse θ;
2. todas as amostras retiradas da população, de acordo com certo procedimento;
3. para cada amostra, calculamos o valor t da estatística T; e
4. os valores t formam uma nova população, cuja distribuição recebe o nome de distribuição amostral de T.



Mas como poderemos pelo menos fazer um histograma de valores da estatística se só dispomos de uma amostra?

Vamos simplificar o problema de estimação de um parâmetro genérico θ para um problema específico de estimação da média populacional, μ.
Para isso dispomos de uma amostra aleatória de tamanho n da população cujos valores observados são $x_1, x_2, \ldots, x_n$.

No que segue usaremos:

μ para a média da população e

- μ para média da população
- σ^2 para variância da população (σ - desvio padrão da população)
- Um estimador natural de μ a ser usado é a média amostral $\overline{X}$.

Teorema Central do Limite (TCL)

Se $X_1, X_2, \ldots, X_n$ é uma amostra aleatória simples de uma população qualquer cuja a média é μ e a variância é σ^2, a distribuição amostral de $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, a média amostral, se aproxima de uma distribuição normal com média μ e a variância $\frac{\sigma^2}{n}$ quando n cresce.

Ou seja, para n suficientemente grande,

$$\overline{X} \utilde{a} N\left(\mu, \frac{\sigma^2}{n}\right)$$

ou equivalentemente,

$$\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \utilde{a} N(0,1)$$

5. Estimação

A Inferência Estatística tem por objetivo fazer generalizações sobre uma população, com base nos dados de uma amostra. Existem dois problemas básicos nesse processo:

(a) estimação de parâmetros e

(b) teste de hipóteses sobre parâmetros.

Lembrem-se que parâmetros sã funções de valores populacionais, enquanto estatísticas são funções de valores amostrais.

Um estimador T do parâmetro θ é qualquer função das observações na amostra, ou seja,

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

O problema de estimação pode ser descrito como o problema de determinar uma função $T = g(X_1, X_2, \cdots, X_n)$ que seja "próxima" de θ, segundo algum critério.

O estimador T do parâmetro θ é não viesado se

$$E[T] = \theta, \quad \forall\theta$$

Observação: Estimativa é o valor assumido pelo estimador em uma particular amostra.

Uma sequência $\{T_n\}$ de estimadores de um parâmetro θ é consistente se para todo $\epsilon > 0$,

$$P(|T_n - \theta| \geq \epsilon) \to 0$$

quando $n \to \infty$.

Proposição: Uma sequência $\{T_n\}$ de estimadores de θ é consistente se

$$\begin{array}{cc} (1) & \lim_{n\to\infty} E[T_n] = \theta \\ (2) & \lim_{n\to\infty} Var[T_n] = 0 \end{array}$$

Se T e T' são dois estimadores não viesados de θ e $Var(T) < Var(T')$, então T é um estimador mais eficiente do que T'

Chama-se erro quadrático médio (EQM) do estimador T do parâmetro θ ao valor

$$EQM(T;\theta) = E[(t-\theta)^2]$$

$$EQM(T;\theta) = Var(T) + (E[T]-\theta)^2$$

Desigualdade de Tchebyshev

> Para provar a consistência de um estimador usa-se a desigualdade de Tchebyshev

Seja X uma variável aleatória com valor esperado $E[X] = \mu$ e variância $Var(X) = \sigma^2$.

Então. para todo $t > 0$,

$$P(\mid X - \mu \mid \geq t) \leq \frac{\sigma^2}{t^2}$$

Se $X_1, X_2, \ldots, X_n$ é uma amostra aleatória de uma população cuja média (valor esperado) é μ e cuja variância é σ^2, vimos que $\overline{X}$ é uma variável aleatória com média μ e variância $\frac{\sigma^2}{n}$.

Usando a desigualdade de Tchebyshev, tem-se, para todo $t > 0$,

$$P(\mid \overline{X} - \mu \mid \geq) \leq \frac{\sigma^2}{n \,.\, t^2}$$

Lei dos Grandes Números

Considere a repetição independente de n ensaios de Bernoulli cuja probabilidade de sucesso é p, $0 < p < 1$, e seja k o número de sucessos nos n ensaios. A Lei dos Grandes Números (LGN) afirma que, para n grande, a proporção observada de sucessos k/n estará próxima de p.

Formalmente, para todo $\epsilon > 0$.

$$P\left(\mid \frac{k}{n} - p \mid \geq \epsilon \right) \leq \frac{p(1-p)}{n\epsilon^2}$$

A demonstração da LGN segue da desigualdade de Tchebyshev.

Métodos de Estimação

1. Métodos dos Momentos

Nesse método as propostas de estimadores são feitas igualando-se os momentos populacionais aos momentos amostrais correspondentes. O momento populacional de ordem k, $k \in \mathbb{N}$ e definido por $M_k = E[X_k]$.

O momento amostral de ordem k, dada uma amostra aleatória $X_1, X_2, \ldots, X_n$ da população é definido por $m_k = \frac{1}{k} \sum_{i=1}^{n} X_i^k$.

Nesse método, para encontrar estimadores de parâmetros, resolvemos equações do tipo

$$m_k = M_k$$

ou seja, usamos os momentos amostrais como estimadores dos momentos populacionais.

2. Método da máxima verossimilhança

O princípio da verossimilhança afirma que devemos escolher aquele valor do parâmetro desconhecido que maximiza a probabilidade de obter a amostra particular observada, ou seja, o valor que torna aquela amostra a “mais provável”.

- Função de verossimilhança

Dada uma amostra aleatória simples de tamanho $n : X_1, X_2, \ldots, X_n$ tem-se n variáveis aleatórias independentes e identicamente distribuídas.

Portanto, a função de densidade de probabilidade (função de probabilidade) conjunta f_n (p_n) fatora nas funções de densidade de probabilidade (funções de probabilidade) marginais.

$$f_n(x_1, x_2, \ldots, x_n; \theta) = f(x_1; \theta) f(x_2; \theta) \ldots f(x_n; \theta)$$

ou

$$p_n(x_1, x_2, \ldots, x_n; \theta) = p(x_1; \theta) p(x_2; \theta) \ldots p(x_n; \theta)$$

em que θ é um parâmetro de caracteriza a distribuição da população. De fato, θ pode representar mais de um parâmetro quando for o caso.

Em geral, θ é desconhecido e, depois de observar a amostra temos os valores $x_1, x_2, \ldots, x_n$.

Podemos então, olhar a densidade (probabilidade) conjunta como uma função de θ.

$$L(\theta; x_1, \ldots, x_n) = p_n(x_1, \ldots, x_n; \theta)$$

tem-se assim a função de verossimilhança.

O estimador de θ é então obtido, maximizando-se a função de verossimilhança.

Observação: Como a função de verossimilhança é não negativa e a função log. natural é estritamente crescente, o máximo da função de verossimilhança ser à equivalente ao máximo a função log. natural da função de verossimilhança.

$$l(\theta; x_1, \ldots, x_n) = ln\{L(\theta; x_1, x_2, \ldots, x_n)\}$$

Propriedades dos estimadores de máxima verossimilhança:

1. Os Estimadores da máxima verossimilhança são constantes
2. Os Estimadores de máxima verossimilhança são invariantes sob transformações: se $\hat{\theta}$ é estimador de máxima verossimilhança de θ, segue que $g(\hat{\theta})$ é o estimador de máxima verossimilhança de $g(\theta)$

6. Intervalos de Confiança

Intervalos de Confiança com nível de confiança γ para a média populacional Amostras da distribuição normal ou amostras suficientemente grandes n ≥ 30

$$IC(\mu, \gamma) : \overline{X} \pm z_{\left(\frac{1+\gamma}{2}\right)} \cdot \frac{\sigma}{\sqrt{n}}$$

Observação: se o valor de σ não for conhecido substitua-o na expressão acima por uma estimativa.

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_1 - \overline{x})^2}$$

Inferência na Normal

1. $\overline{X} \sim N(\mu; \frac{\sigma^2}{n})$
2. $S^2 = \frac{\sum(x_i - \overline{x})^2}{n-1}$
3. $(\frac{n-1}{\sigma^2})S^2$
4. $T = \frac{\overline{X}-\mu}{S/\sqrt{n}}$

Intervalo de Confiança (IC)	Z
80	1,28
85	1,44
90	1,64
95	1,96
99	2,57
99,5	2,80
99,9	3,29

Intervalo de confiança para a proporção amostral

Nesse caso, a população (X) é considerada uma Bernoulli (p) em que p é à proporção populacional que desejamos estimar.

$$IC(p, \gamma) : \hat{p} \pm z_{(\frac{1+\gamma}{2})} \cdot \left(\sqrt{\frac{1}{4n}} ou \frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}\right)$$

Quando a Variância é desconhecida

$$IC(\mu, \gamma) : \overline{X} \pm t_{(\frac{1-\gamma}{2}, n-1)} \cdot \frac{s}{\sqrt{n}}$$

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_1 - \overline{x})^2}$$

No R:

```r
x <- c(82, 102, 91, 90, 87, 107, 83, 78, 88, 101, 99, 76, 67,
87, 99, 88) #Conjunto de Amostra
desvioAmostral <- function(x,n){
    desvio <- 0
    for (i in x){
        if (mean(x)>=i){
            desvio <- desvio + ((mean(x)-i)^2)
        }
        if(mean(x)<=i){
            desvio <- desvio + ((i-mean(x))^2)
        }
    }
    desvio <- desvio / (n-1)
    return(sqrt(desvio))
}
desvioAmostral(x,length(x))
```

7. Testes de Hipóteses

> Uma **hipótese** é uma afirmativa sobre um parâmetro , ou seja, sobre uma característica da população.

Um **teste de hipótese** é um procedimento para testar uma hipótese baseado numa amostra da população

Regra do evento raro: Se, sob uma suposição, a probabilidade de um evento particular observado é excepcionalmente pequena, concluímos que a suposição provavelmente não está correta.

Exemplo: Se um produto que permite escolher o sexo de uma criança for testado por 100 casais. Podemos obter 2 resultados. (a) 52 meninas e (b) 97 meninas. Embora ambos estejam “acima da média” (50), o resultado 52 não é significativo enquanto que o 97 é um resultado significativo.

Fundamentos do Teste de Hipótese

1. Hipótese Nula (H_0) e Alternativa (H_1)

A hipótese nula, denotada por H_0, é uma afirmativa sobre um parâmetro. Por exemplo: μ = 90, p=0,10, σ ≥ 2 etc. A hipótese alternativa, denotada por H_1, é uma afirmativa complementar à hipótese nula tal que não exista interseção entre as duas hipóteses. Por exemplo: μ > 90, p $\neq$ 0, 10, σ < 2 etc.

Temos que decidir por uma das duas hipóteses baseando-nos numa amostra da população. Logo, estamos sujeitos a dois erros diferentes.

Decisão	H_0 é Verdadeira	H_0 é Falsa
Assumir H_0 como Falsa	Erro Tipo I	sem erro
Assumir H_0 como Verdadeira	sem erro	Erro Tipo II

1. **Estatística de Teste: é** uma função que produz um valor real com base nos dados amostrais.

2. **Região Crítica:** Uma regra de decisão ou procedimento de teste consiste em especificar um conjunto de valores da estatística de teste para os quais rejeitaremos a hipótese nula (H_0). Chamamos esse conjunto de valores, para os quais rejeitaremos H_0, de Região Crítica do teste.

3. **Nível de Significância (α) do teste:** é a probabilidade de ser cometer o erro do tipo I, ou seja, é a probabilidade de rejeitar uma hipótese nula verdadeira.

   Quando não é mencionado adota-se α = 5%. Os valores comuns para α são 10% 5% e 1%

α	Z_c
5%	1,645
10%	1,28
1%	2,32

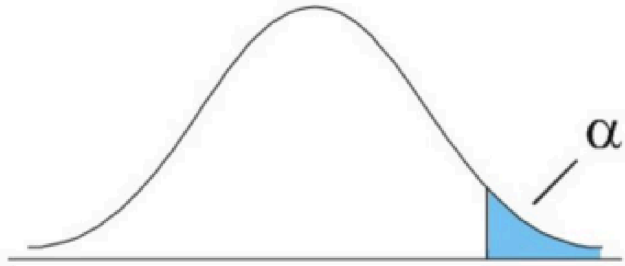
4. **Erro do Tipo II**: usamos a letra grega β para representar a probabilidade de cometer o erro tipo II: “não rejeitar uma hipótese nula falsa”.

5. **Testes Bilaterais e Unilaterais:**

Unilateral à esquerda:

H_0: $\mu = 50$

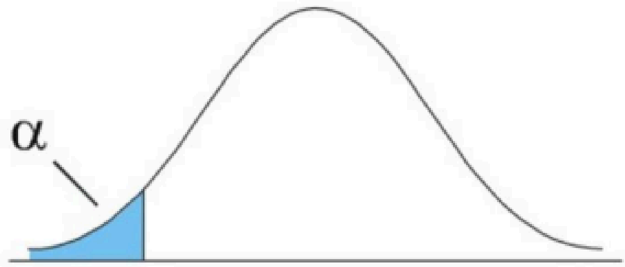
H_1:: $\mu > 50$



Unilateral à direita:

H_0: : $\mu = 50$

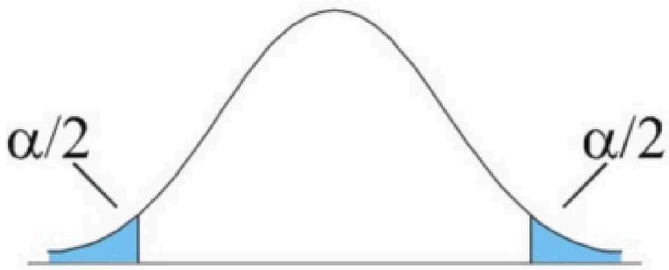
H_1: : $\mu < 50$



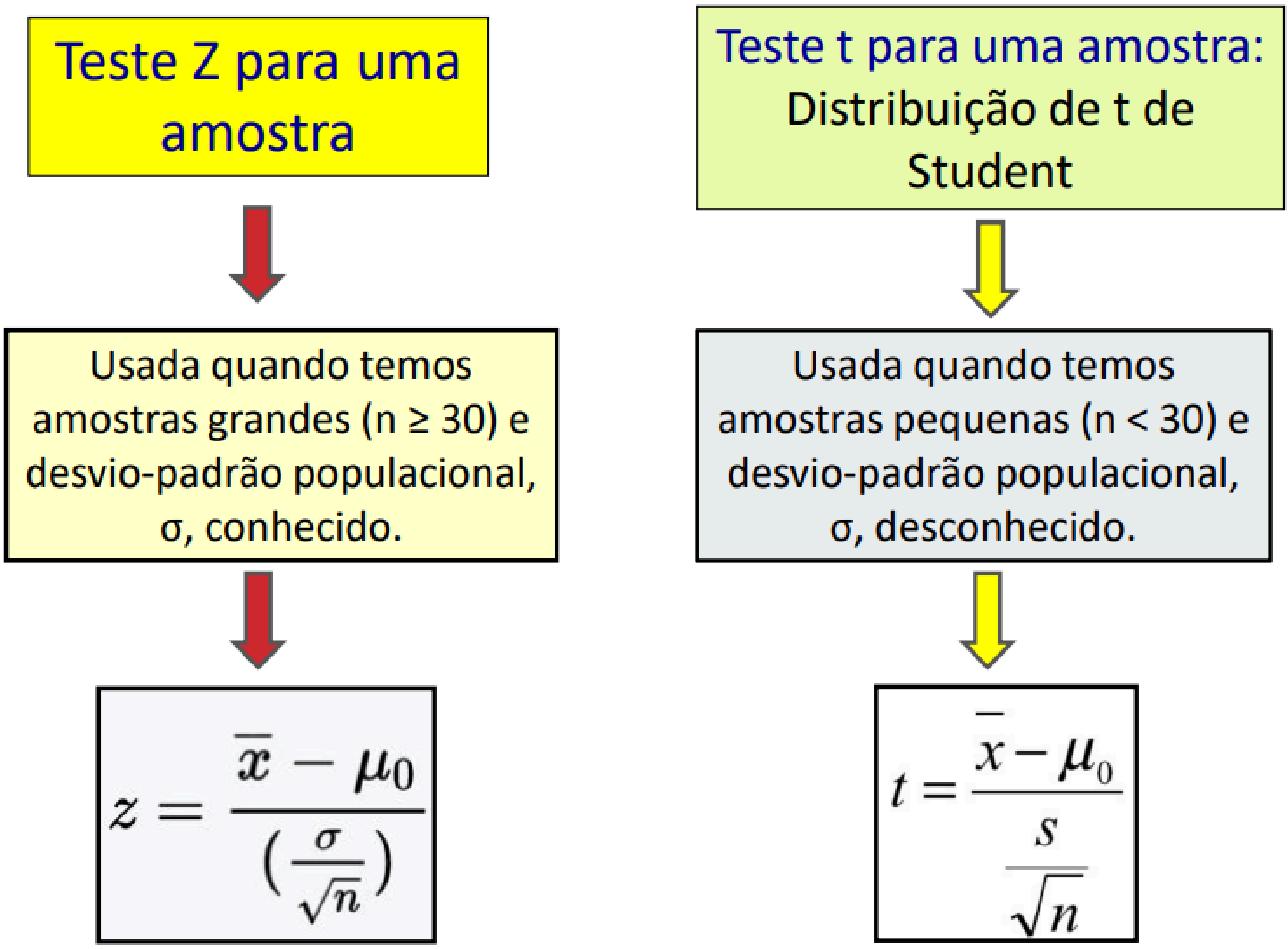
Bilateral:

H_0: : $\mu = 50$

H_1:: $\mu \neq 50$



1. Procedimento clássico de testes de hipóteses
   - Passo 1: Fixe a hipótese nula a ser testada e qual é a forma da hipótese alternativa.
   - Passo 2: Use a teoria estatística e as informações disponíveis para decidir qual estatística seŕ a usada no teste. Obtenha a distribuição amostral da estatística de teste.
   - Passo 3: Fixe o nível de significância α do teste, isto é, a probabilidade de rejeitar uma hipótese nula verdadeira e determine a região crítica do teste.
   - Passo 4: Use a amostra para calcular o valor amostral da estatística de teste.
   - Passo 5: Se o valor amostral cair na região crítica, rejeite H_0, caso contrário, não rejeite H_0.



Para n grande:

$$\frac{\hat{p} - p}{\sqrt{p(1-p)/n}}$$

1. **p-Valor ou Nível Descritivo ou Probabilidade de significância**

   Outra maneira de proceder consiste em apresentar o p-valor do teste. De maneira informal, o p-valor caracteriza o grau de adesão dos dados amostrais à hipótese nula. É calculado usando-se uma probabilidade condicional, supondo que H_0 é verdadeira. Portanto, o p-valor está entre 0 e 1. Na prática, rejeitaremos H_0 para p-valores muito pequenos.

p-valor	Natureza da evidência contra H_0
0,10	Marginal
0,05	Moderada
0,025	Substancial
0,01	Forte
0,005	Muito Forte
0,001	Fortíssima