Inferência estatística

Processo de tirar conclusões sobre um conjunto maior (população) usando informação de um conjunto menor (amostra).

- População: todos os casos ou situações sobre as quais o pesquisador quer fazer inferências.
  - fazer inferências sobre concentração de poluentes num determinado lençol freático,
  - predizer a quantidade de petróleo num poço a ser perfurado,
  - estimar o tempo de vida útil de um componente eletrônico.
- Amostra: um subconjunto qualquer da população. Por que não observar a população inteira?
  - Alto custo.
  - Tempo muito longo.
  - Impossibilidade física (e.g. estudo de poluição amosférica).
  - Impossível lógica (e.g. em ensaios destrutivos).
- Variáveis: características de uma população que diferem de um indivíduo para outro e as quais queremos estudar.
- Observações: medidas de uma ou mais variáveis de um indivíduo na amostra.

Princípios de estimação

Estamos interessados em um parâmetro populacional $\theta \in \Theta$.

Θ: espaço paramétrico.

- Se $X \sim Poisson(\theta)$, então $\Theta = \{\theta : \theta > 0\}$.
- Se $X \sim N(\mu, 1)$, então $\Theta = \{\mu : -\infty < \mu < \infty\}$.
- Se $X \sim N(\mu, \sigma^2)$, então $\Theta = \{(\mu, \sigma^2) : -\infty < \mu < \infty,\ \sigma^2 > 0\}$.

Qual o valor mais plausível de θ com base nos dados amostrais?

$$\boxed{\text{estimativa pontual de } \theta}$$

Exemplos: média amostral, desvio padrão amostral, etc.

Definição: Uma estatística é uma função qualquer dos elementos da amostra e que não depende do parâmetro desconhecido.

Notação geral,

- Estatísticas: letras latinas, e.g. $\overline{x}$ (média amostral), s (desvio padrão amostral), etc.
- Parâmetros: letras gregas, e.g. μ (média populacional), σ (desvio padrão populacional).

Conforme a amostra aumenta, mais informação teremos sobre a população e mais precisas serão as estimativas.

Definição: Qualquer estatística que assume valores em Θ é denominada um estimador para θ.

Qualquer estimador é uma estatística mas nem toda estatística define um estimador.

Exemplo: 10 repetições de um ensaio de Bernoulli,

$$X_i = \begin{cases} 1, & \text{se ocorre sucesso} \\ 0, & \text{se ocorre fracasso} \end{cases}$$

Parâmetro desconhecido: probabilidade de sucesso p.

$Y = \sum_{i=1}^{n} X_i$ é uma estatística porém não é um estimador de p.

Um possível estimador para p seria

$$\hat{p} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{Y}{n}$$

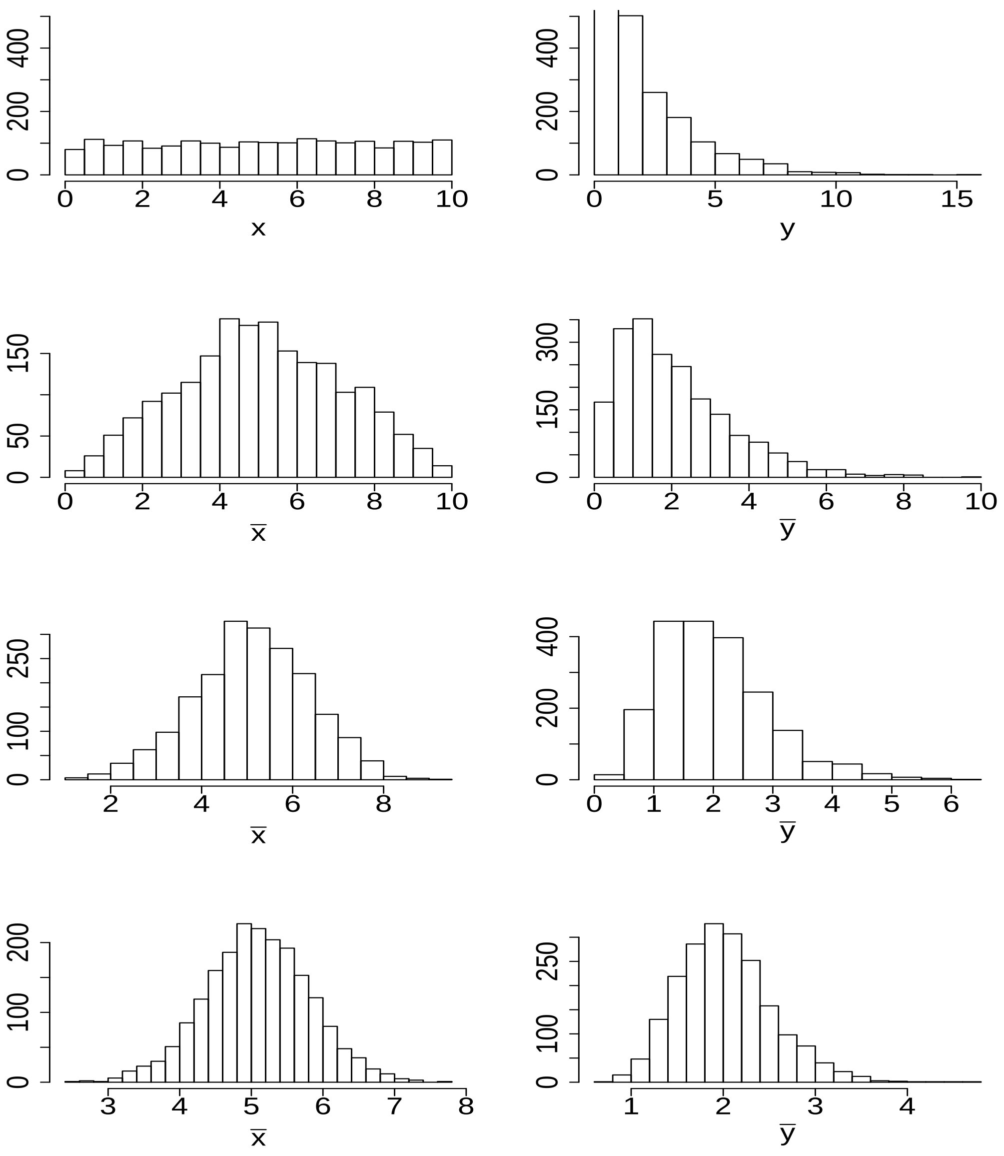
Se foram obtidos 3 sucessos então $\hat{p} = 0,3$ é uma estimativa de p.

Y/n é uma v.a. com possíveis valores $0, 1/10, \ldots, 1$.

Vamos assumir que dispomos de uma amostra segundo a definição a seguir.

Definição: Se X representa uma característica de interesse da população, uma amostra aleatória de tamanho n é o conjunto de n variáveis aleatórias independentes $X_1, \ldots, X_n$ cada uma com a mesma distribuição de X.

Teorema Central do Limite



Para amostras grandes $X_1, \ldots, X_n$, se

$$E(X_i) = \mu \qquad \text{e} \qquad Var(X_i) = \sigma^2$$

para $i = 1, \ldots, n$ temos o seguinte resultado

A distribuição da média amostral $\overline{X}$ é aproximadamente Normal com média μ e desvio padrão $\sigma/\sqrt{n}$.

Usaremos a notação

$$\overline{X} \sim N(\mu, \sigma^2/n).$$

Testes de Hipóteses (Um exemplo)

- Experimento: Teste do tipo certo-errado com 10 questões.
- Objetivo: Testar se o aluno está advinhando.
- Seja a v.a. X="número de acertos em 10 questões".
- Suposição: $X \sim Binomial(10, p)$.
- Queremos testar,

$$H_0 : p = 1/2 \quad \times \quad H_1 : p > 1/2.$$

- Regra de decisão: "o aluno não está advinhando se acertar 8 ou mais questões. " (Rejeitar H_0 se $X \geq 8$).
- $$P(X \geq 8 \,|\, p = 1/2) = \sum_{k=8}^{10} \binom{10}{k} 0,5^{10} = \frac{7}{128} \approx 0,054.$$

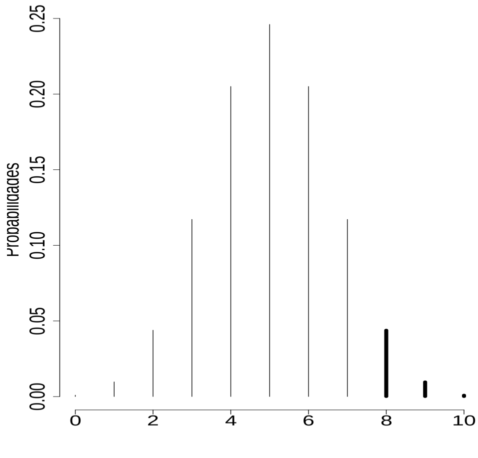
Testes de Hipóteses (Outro exemplo)

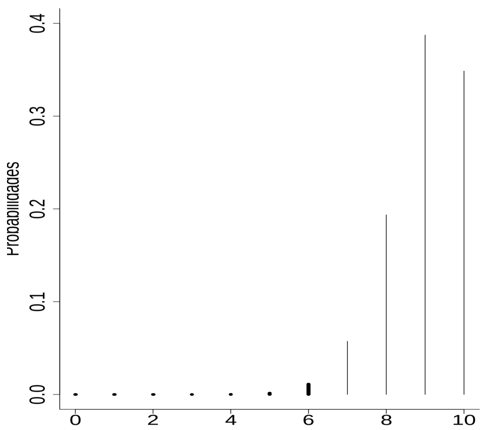
- Afirmação: Fornecedor garante que 90% de sua produção não apresenta defeito.
- Experimento: Selecionar ao acaso 10 itens de um lote e contar o número de defeituosos.
- Regra de decisão: não comprar o lote se o número observado de não defeituosos for muito pequeno.
- X="número de não defeituosos na amostra de 10 itens".
- Suposição: $X \sim Binomial(10, p)$.
- Queremos testar,

$$H_0 : p = 0,9 \quad \times \quad H_1 : p < 0,9.$$

Qual o valor de k tal que $P(X \leq k \mid p = 0,9) < 0,025$?

$$\begin{aligned} P(X \leq 5 \mid p = 0,90) &= 0,001 \\ P(X \leq 6 \mid p = 0,90) &= 0,012 \\ P(X \leq 7 \mid p = 0,90) &= 0,069. \end{aligned}$$


Probabilidades
0.00
0.05
0.10
0.15
0.20
0.25
0
2
4
6
8
10


Probabilidades
0.0
0.1
0.2
0.3
0.4
0
2
4
6
8
10

Tipos de erro

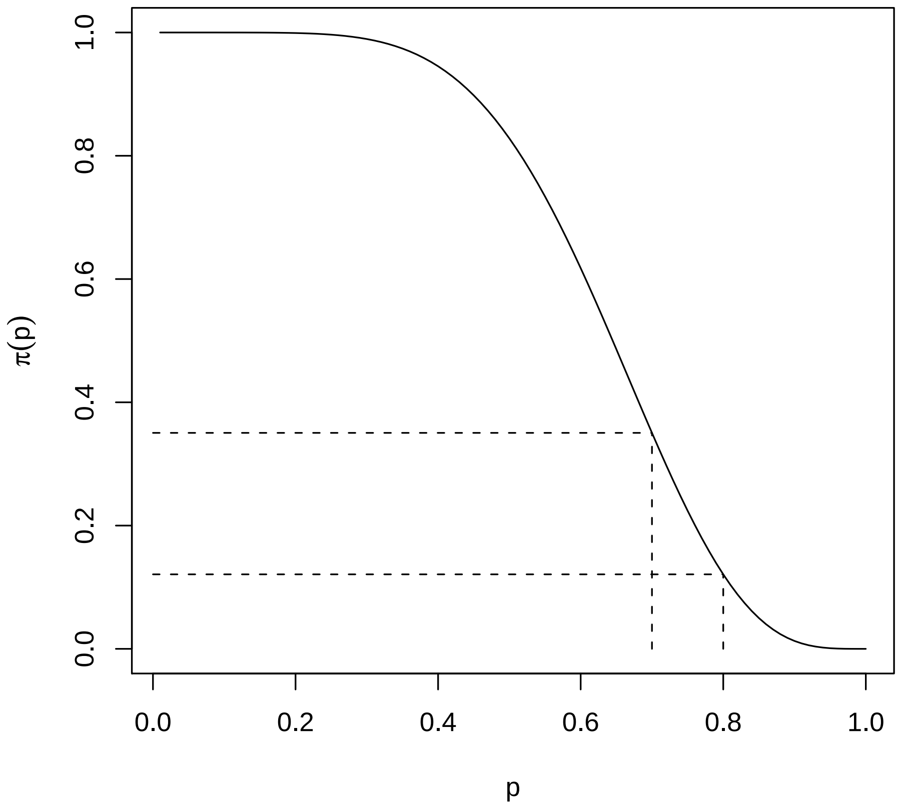
$P(\text{erro tipo I }) = P(\text{rejeitar } H_0 \mid H_0 \text{ verdadeira}) = \alpha$

$P(\text{erro tipo II}) = P(\text{não rejeitar } H_0 \mid H_0 \text{ falsa}) = \beta$

Função poder: $P(\text{rejeitar } H_0 \mid \theta)$

Não é possível minimizar α e β simultaneamente

Na prática é costume fixar um valor (pequeno) para α.



1.0
0.8
0.6
0.4
0.2
0.0
π(p)
0.0
0.2
0.4
0.6
0.8
1.0
p

	Decisão	
Verdade	Aceitar H_0	Rejeitar H_0
H_0 verdadeira	Decisão correta (probabilidade $1-\alpha$)	Erro Tipo I (probabilidade α)
H_0 falsa	Erro Tipo II (probabilidade β)	Decisão correta (probabilidade $1-\beta$)

Nível Descritivo (p-valor)

- No teste com 10 questões suponha que $x_{obs} = 9$.

  $P(X \geq 9 \,|\, p = 1/2) = \binom{10}{9} 0,5^{10} + \binom{10}{10} 0,5^{10} = 0,0107.$

  Decisão: rejeitar H_0 $\forall$ $\alpha > 0,0107$.

α	Decisão
0,050	rejeitar H_0
0,025	rejeitar H_0
0,010	aceitar H_0

- No exemplo do fornecedor suponha que $x_{obs} = 4$.

$$P(X \leq 4 \mid p = 0,90) = 0,000146.$$

  Decisão: rejeitar H_0 $\forall$ $\alpha > 0,000146$.

α	Decisão
0,050	rejeitar H_0
0,025	rejeitar H_0
0,010	rejeitar H_0
0,001	rejeitar H_0