Teste de Hipótese

- Motivação: Um algoritmo que desenvolvido por uma empresa de software retorna o resultado de interesse (dentro do tempo desejado) em 60% dos casos em que é utilizado. Com o objetivo de incrementar essa porcentagem, a empresa desenvolveu um novo algoritmo elaborado com técnicas avançadas e afirma: "O novo algoritmo é melhor do que o anterior".

Teste de Hipótese

- Se o novo algoritmo produz o efeito desejado (retorna o resultado de interesse dentro do tempo desejado) em 90% dos casos (em uma amostra) nos quais foi utilizado:
  - A afirmação da empresa tem fundamento?
  - Para decidir se a afirmação tem fundamento, baseamos nosso estudo na informação contida numa amostra aleatória de execuções do algoritmo.
  - Vamos supor que o algoritmo novo não pode ser pior do que o anterior.

Teste de Hipótese

- Decisão: Estratégia de Decisão
  - Seja θ o parâmetro ou a quantidade de interesse (proporção de vezes, our probabilidade, do algoritmo retornar o resultado de interesse dentro do tempo desejado).
  - $\theta \in \omega$ (espaço paramétrico)
  - Suponha que $\omega \doteq \omega_1 \cup \omega_2$ (união disjunta)
  - Teste: $H_0 : \theta \in \omega_1 \quad vs \quad H_1 : \theta \in \omega_2$.

Teste de Hipótese

- Convenção:
  - H_0: hipótese nula: a afirmação referida ao possível valor de θ que, algumas (muitas) vezes, entende-se não ser verdadeira .
  - H_1: hipótese alternativa: a afirmação referida ao possível valor de θ, que algumas (muitas) vezes, entende-se que seja verdadeira.

Teste de Hipótese

- No exemplo de motivação: $H_0 : \theta = 0.6 \quad vs \quad H_1 : \theta = 0.9$
- Definimos o conjunto de decisões $A = \{a : a = a_1 \text{ ou } a = a_2\}$:
  - a_1 = aceitar H_0 (e rejeitar H_1)
  - a_2 = rejeitar H_0 (e aceitar H_1)
  - a $(= a_1$ ou $= a_2)$ será a decisão tomada a partir da observação da amostra.

Regra de Decisão

- A amostra $\mathbf{X} = (X_1, ..., X_n)^t$ pertence ao espaço amostral S.
- S pode ser definido como $S = S_1 \uplus S_2$ de forma que $\mathbf{X} \in S_1$ ou $\mathbf{X} \in S_2$
- Regra:
    - se $\mathbf{X} \in S_1$, então $a = a_1$ (aceita-se H_0)
    - se $\mathbf{X} \in S_2$, então $a = a_2$ (rejeita-se H_0)

Regra de Decisão

- Exemplo: Seja X uma observação proveniente da Normal $N(\mu, 1)$, μ desconhecido.
  - Teste: $H_0 : \mu = 0 \quad vs \quad H_1 : \mu \neq 0$
    - se $X > \frac{1}{2}$ ou $X < -\frac{1}{2}$ (S_2), então $a = a_2$ (rejeita-se H_0)
    - se $X \in \left[-\frac{1}{2}, \frac{1}{2}\right]$, então $a = a_1$ (aceita-se H_0)
  - S_2 é chamada de região de rejeição de H_0, ou região crítica do teste.
  - S_1 é chamada de região de aceitação de H_0.

Tipos de erro

- Erro do tipo I $= \{$rejeitar $H_0 | H_0$ é verdadeira$\}$
- Erro do tipo II $= \{$aceitar $H_0 | H_0$ é falsa$\}$

	a_1 (aceitar H_0)	a_2 (rejeitar H_0)
H_0 verdadeira	—	erro tipo I
H_0 falsa	erro tipo II	—

Risco de Decisão

- Erros no exemplo de motivação:

  $H_0 : \theta = 0.6 \quad vs \quad H_1 : \theta = 0.9$

  - Erro tipo I $=$ {rejeitar $H_0|H_0$ é verdadeira} $= \{\theta = 0.9|\theta = 0.6\}$
  - Erro tipo II $=$ {aceitar $H_0|H_0$ é falsa} $= \{\theta = 0.6|\theta = 0.9\}$

- Deseja-se controlar: P(erro tipo I) e P(erro tipo II)

Risco de Decisão

- Montando o teste no exemplo de motivação:

  $H_0 : \theta = 0.6 \quad vs \quad H_1 : \theta = 0.9$

- Seja X : o número de execuções nas quais o algoritmo retornou o resultado esperado dentro do tempo desejado.
- Seja $n = 10$ o tamanho da amostra.
- $X \sim \text{binomial}(10, \theta)$.

Risco de Decisão

- Se H_0 é verdadeira: $X \sim \text{binomial}(10, 0.6)$; $E(X) = 6$
- Se H_0 é falsa: $X \sim \text{binomial}(10, 0.9)$; $E(X) = 9$
- Vamos considerar a seguinte região de rejeição: $S_2 = \{8, 9, 10\}$
- Lembrando que X pode assumir os valores $\{0, 1, 2, ..., 10\}$
- A região crítica S_2 determina as probabilidades de cometer os erros do tipo I e II.

Risco de Decisão

$$
\begin{aligned}
P(\text{erro tipo I}) &= P(\text{rejeitar } H_0 | H_0 \text{ é verdadeira}) \\
&= P(X \in S_2 | \theta = 0.6) \\
&= P(8 \leq X \leq 10 | \theta = 0.6) \\
&= \sum_{x=8}^{10} P(X = x | \theta = 0.6) \\
&\approx 0.1672
\end{aligned}
$$

No R $pbinom(10, 10, 0.6) - pbinom(7, 10, 0.6)$.

Risco de Decisão

$$
\begin{aligned}
P(\text{erro tipo II}) &= P(\text{aceitar } H_0 | H_0 \text{ é falsa}) \\
&= P(0 \leq X \leq 7 | \theta = 0.9) \\
&= \sum_{x=0}^{7} P(X = x | \theta = 0.9) \\
&\approx 0.0702
\end{aligned}
$$

No R $pbinom(7, 10, 0.9)$.

Risco de Decisão

- Observando a relação entre os erros tipo I e II, e S_2:

$H_0 : \theta = 0.6 \quad vs \quad H_1 : \theta = 0.9$

S_2	P(erro tipo I)	P(erro tipo II)
$\{8, 9, 10\}$	0.1672	0.0702
$\{9, 10\}$	0.0463	0.2639
$\{10\}$	0.0060	0.6513

Risco de Decisão

- À medida que diminuímos a probabilidade de cometer o erro do tipo I, diminuímos a região S_2, aumentando a probabilidade de cometer o erro do tipo II.
- Optamos então por controlar o erro tipo I (considerado mais grave, mas isso depende da situação).
- O que seria menos pior: afirmar que o algoritmo novo é melhor, quando possui o mesmo desempenho que o antigo ou afirmar que possui o mesmo desempenho quando, na verdade, é melhor?