Testes de hipóteses

Testes de hipóteses

- Também:
  - Testes estatísticos
  - Testes de significância
- Ferramenta MUITO utilizada na ciência de forma geral
- Obter evidência sobre hipóteses

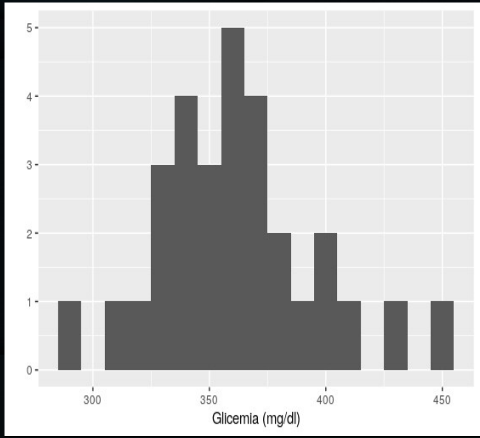
Testes de hipóteses

Exemplo 1

- Como provar que os níveis de glicemia de cães diabéticos são maiores do que o normal?

Exemplo 1

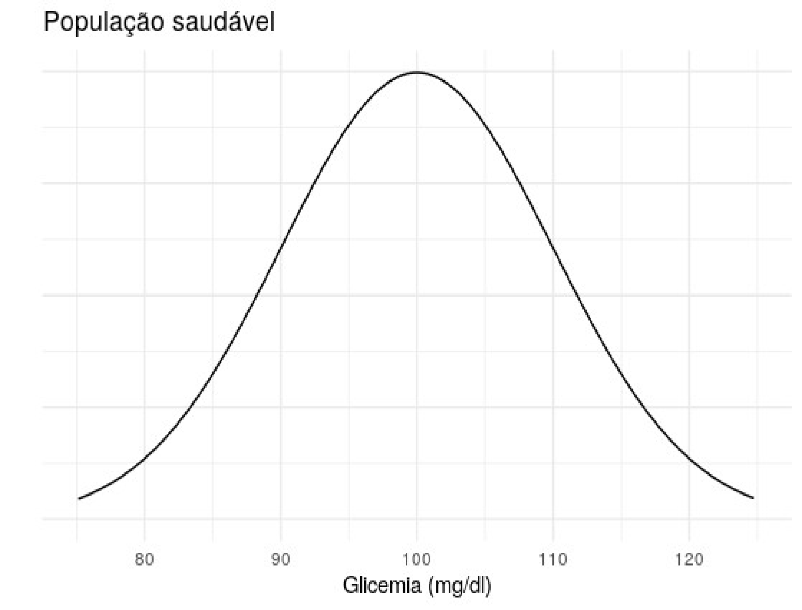
- Como provar que os níveis de glicemia de cães diabéticos são maiores do que o normal?



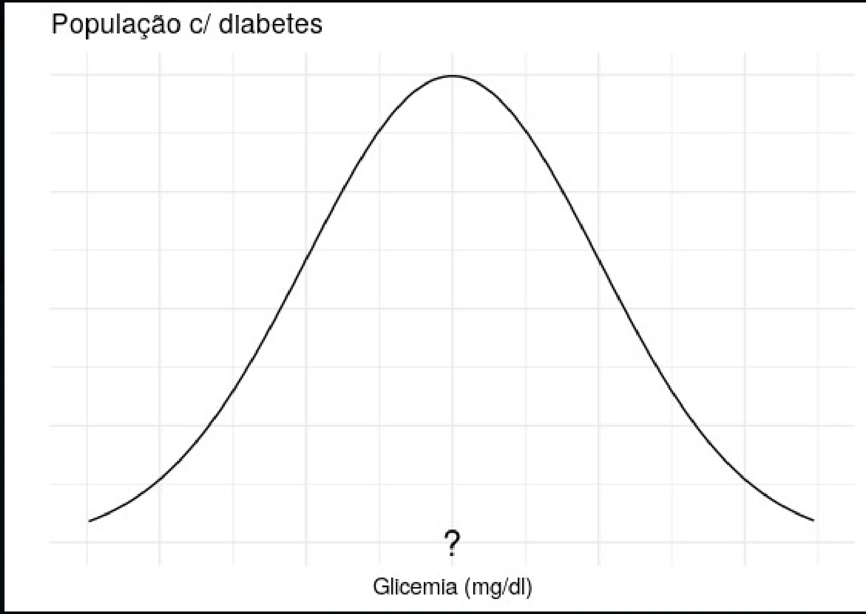
Amostra de cães diabéticos (n = 30)
$\bar{x}$ = 360 mg/dl
s = 34 mg/dl

Exemplo 1

População saudável
μ = 100mg/dl
σ = 10 mg/dl

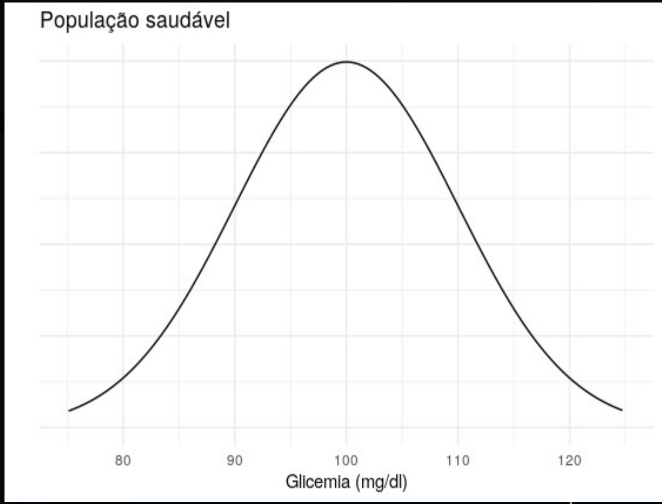


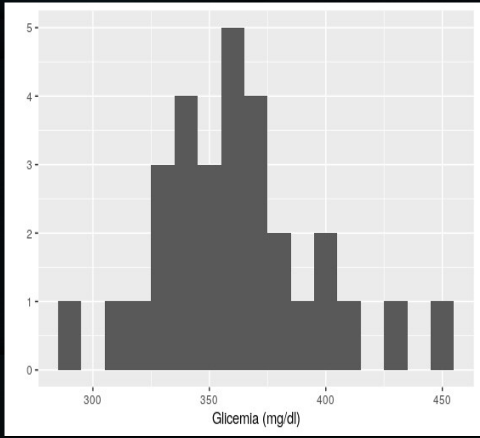
População com diabetes
μ = ? mg/dl
σ = ? mg/dl



Exemplo 1

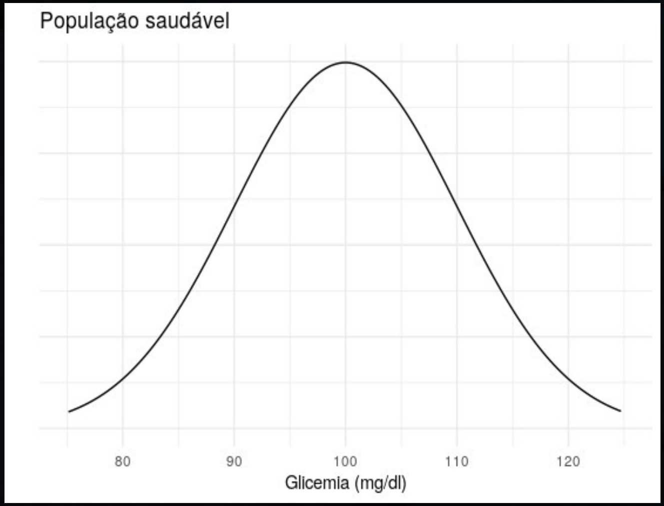
População saudável
μ = 100mg/dl
σ = 10 mg/dl



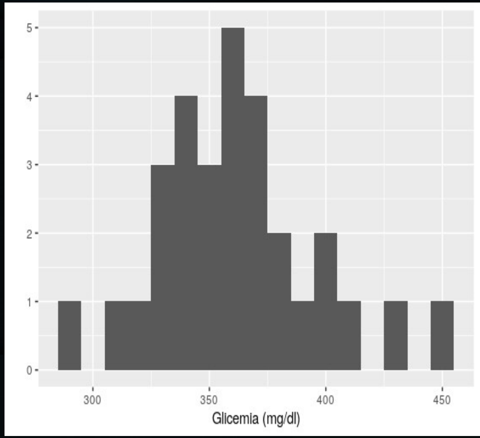


Amostra de cães diabéticos (n = 30)
$\bar{x}$ = 360 mg/dl
s = 34 mg/dl

Exemplo 1



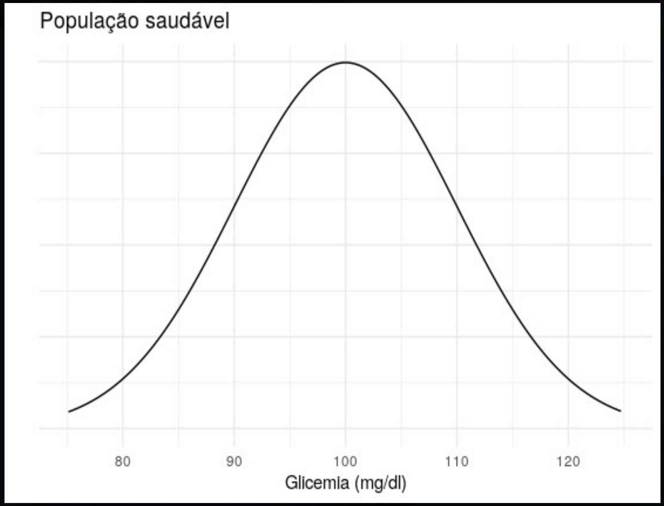
População
μ = 100mg/dl
σ = 10 mg/dl



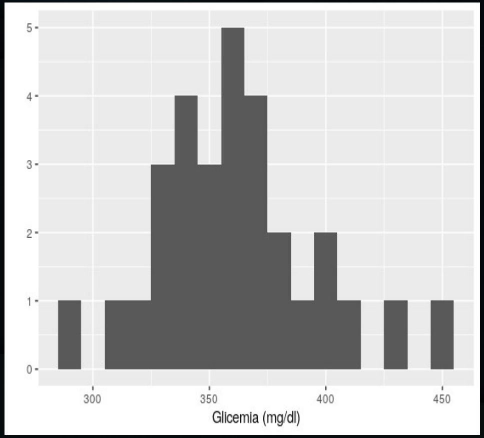
É possível que uma amostra dessa População tenha $\bar{x}$ = 360 mg/dl ?

Amostra de cães diabéticos (n = 30)
$\bar{x}$ = 360 mg/dl
s = 34 mg/dl

Exemplo 1



População
μ = 100mg/dl
σ = 10 mg/dl



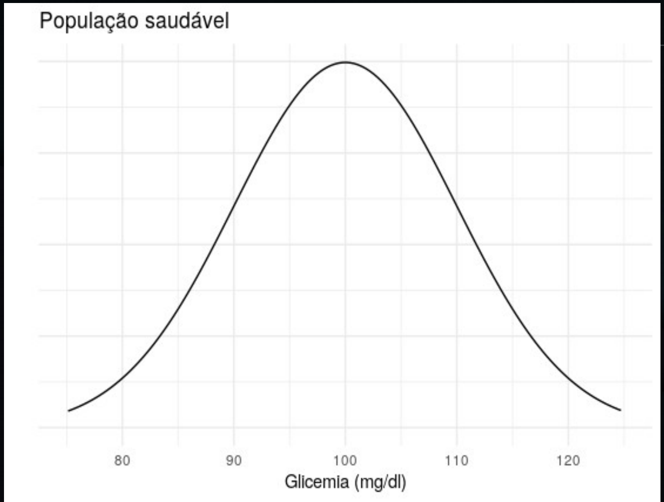
Amostra de cães diabéticos (n = 30)
$\bar{x}$ = 360 mg/dl
s = 34 mg/dl

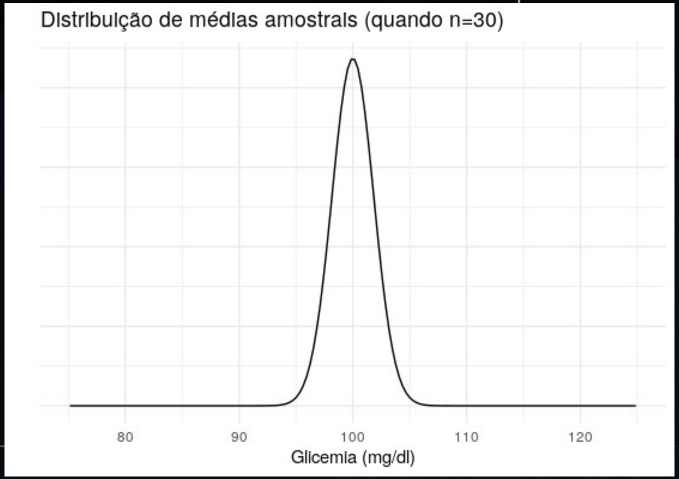
É possível que uma amostra dessa
População tenha $\bar{x}$ = 360 mg/dl ?
Teoricamente, sim, é possível.
Mas é provável?
Não parece…
Pelo TLC eu consigo calcular.

Exemplo 1

População
μ = 100mg/dl
σ = 10 mg/dl

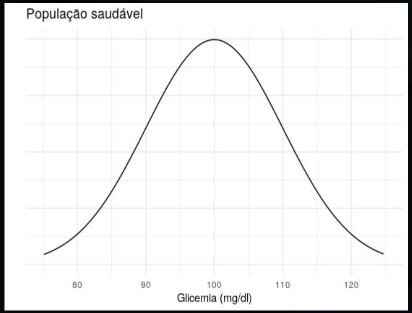
TLC

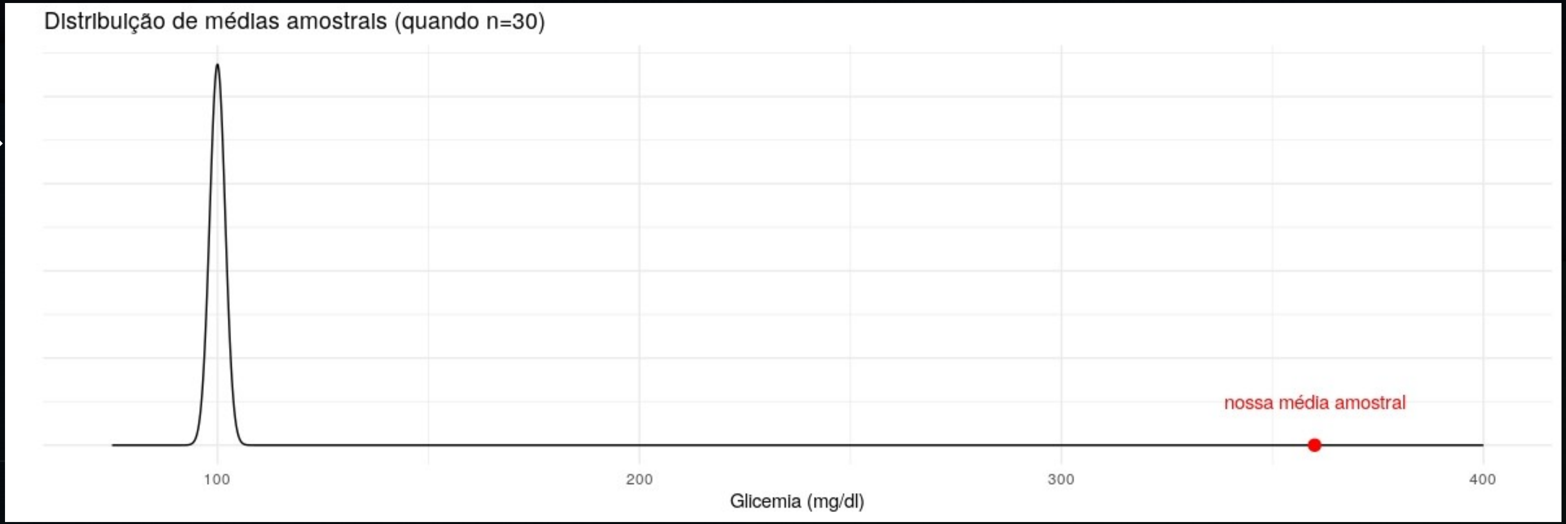




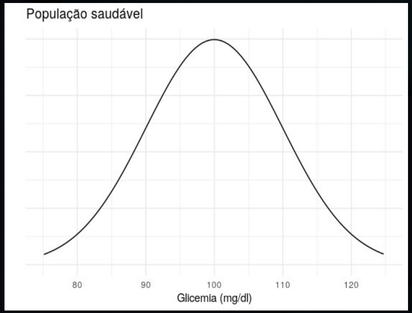
Exemplo 1

TLC



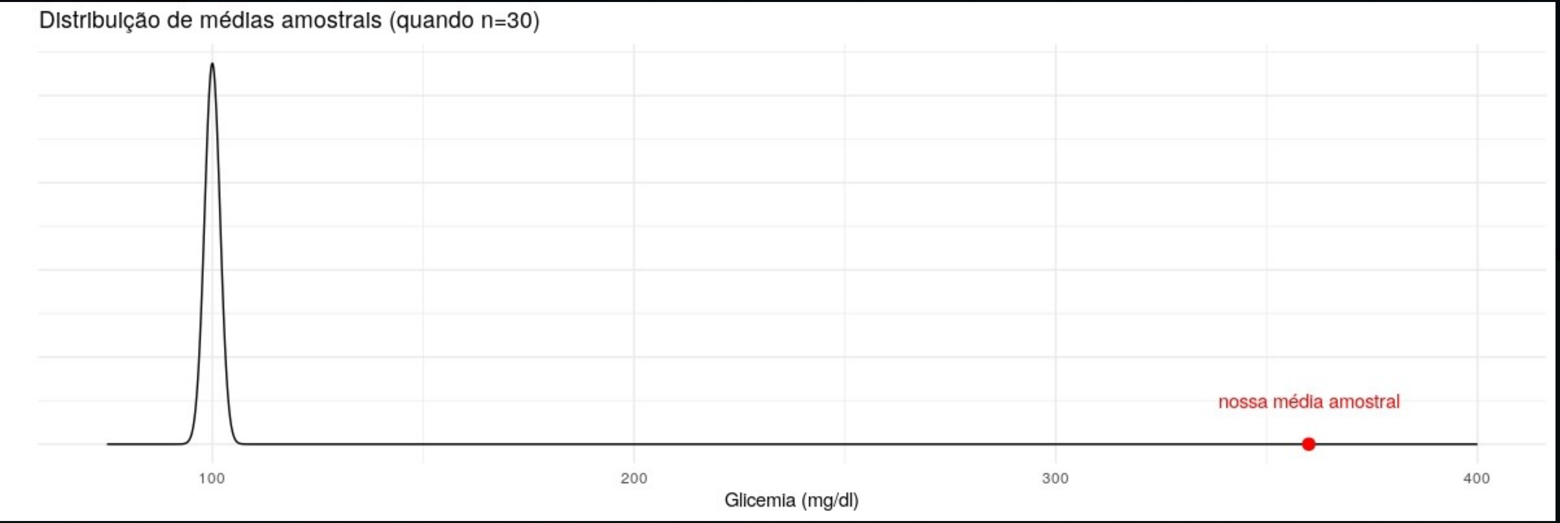


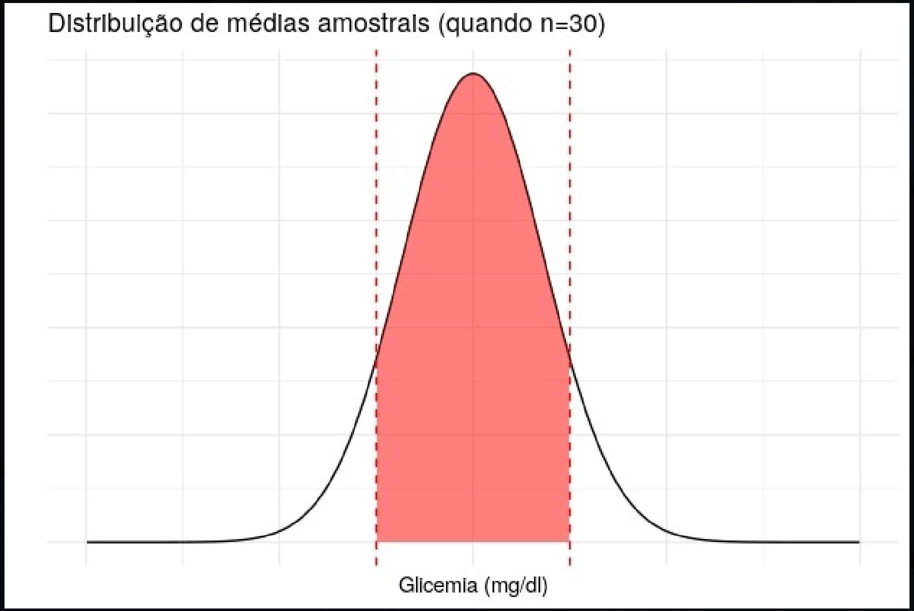
Exemplo 1



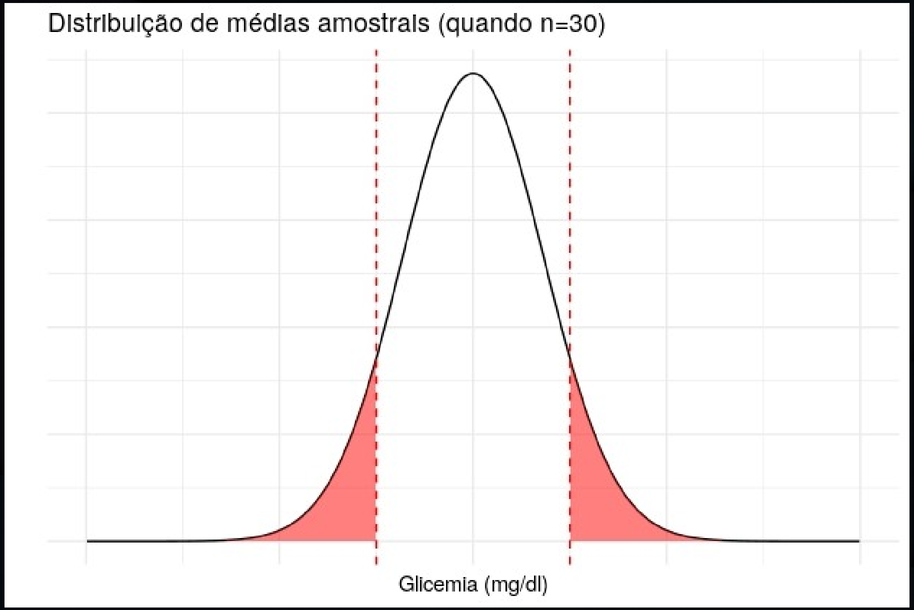
Teoricamente possível, mas muito pouco provável. Quanto?

TLC





Área embaixo da curva:
Probabilidade de uma média amostral ($\bar{x}$) estar nessa faixa de valores

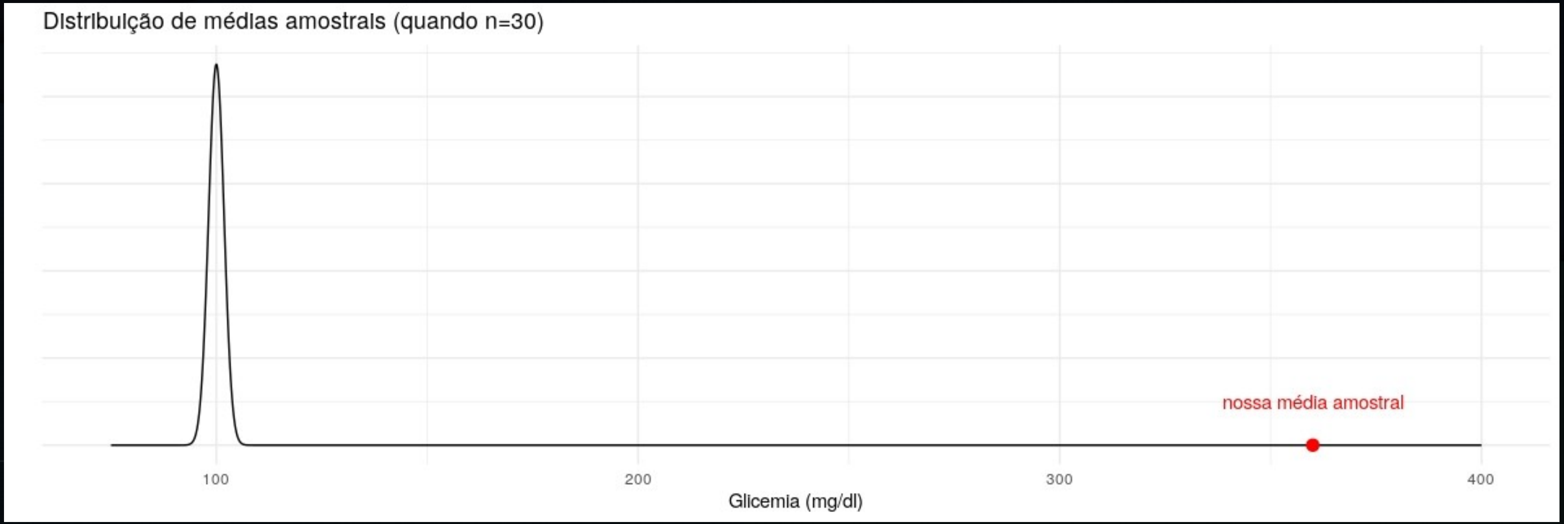


Nesse caso:
Probabilidade de uma $\bar{x}$ ser tão (ou mais) diferente de μ

Área embaixo da curva:
Probabilidade de uma média amostral ($\bar{x}$) estar nessa faixa de valores

Teoricamente possível, mas muito pouco provável.
Quanto? Posso usar o TLC e calcular pela normal.
Mas posso usar também a distribuição t de Student(*), que é mais fácil e não precisa de σ.

Resultado:
P-valor: $2{,}2 * 10^{-16}$ (Isso é 0,00000000000000022)



(*): População deve ter distribuição normal

Formulação de hipóteses

H_0: $\mu = 100$ mg/dl

H_1: $\mu \neq 100$ mg/dl

Procuramos evidência “suficiente” para rejeitar H_0.

Duas maneiras de rejeitar (ou não) H_0

Caso o IC **não inclua** o valor estipulado em H_0

P-valor < α

Rejeita H0

“Tenho evidência de que H_0 é falsa”

Caso o IC **inclua** o valor estipulado em H_0

P-valor >= α

Não rejeita H0

“Não tenho evidência de que H_0 é falsa”

Duas maneiras de rejeitar (ou não) H_0

H_0: $\mu = 100$ mg/dl

H_1: $\mu \neq 100$ mg/dl

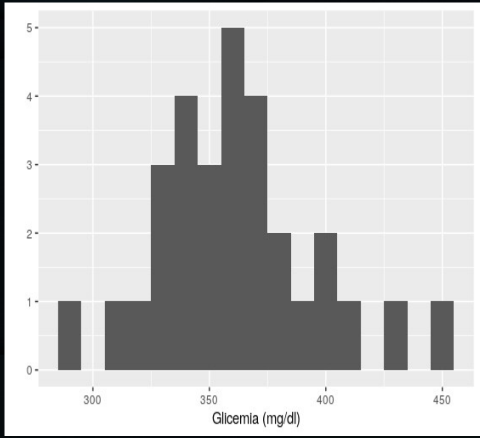
IC(μ,95%): 348,6 - 373,9

Não inclui 100 mg/dl

P-valor: $2,2 * 10^{-16}$

P-valor < 5%

Rejeita H_0



Amostra de cães diabéticos (n = 30)

$\bar{x}$ = 360 mg/dl

s = 34 mg/dl

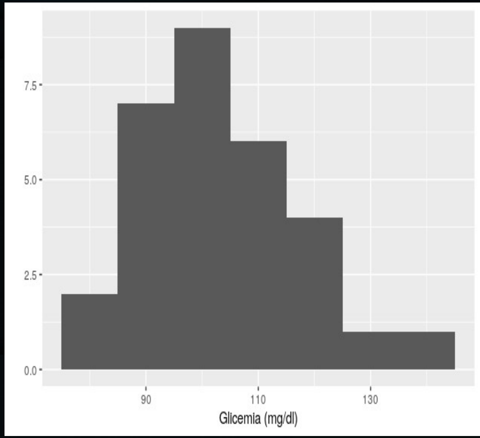
"Tenho evidência de que H_0 é falsa"

Exemplo 2

- Um veterinário está suspeitando que cães com parvovirose também têm níveis de glicemia alterados.

Exemplo 2

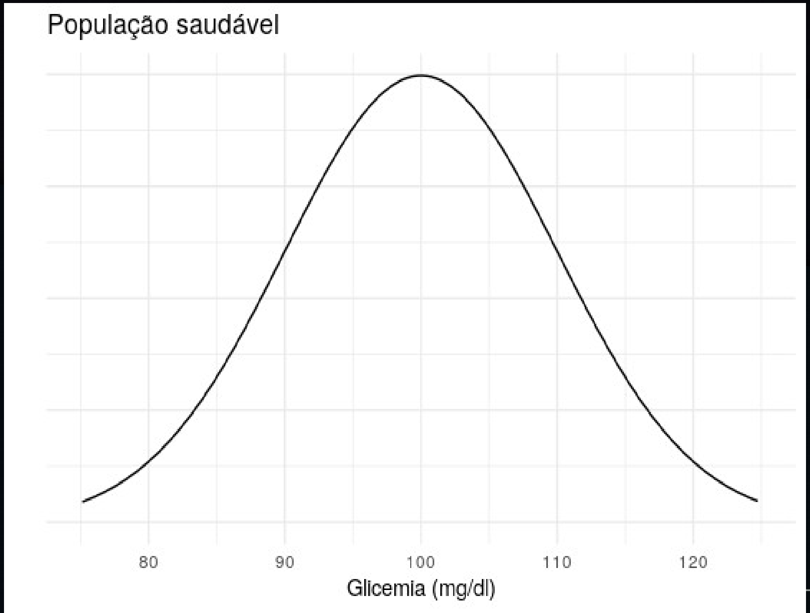
- Um veterinário está suspeitando que cães com parvovirose também têm níveis de glicemia alterados.



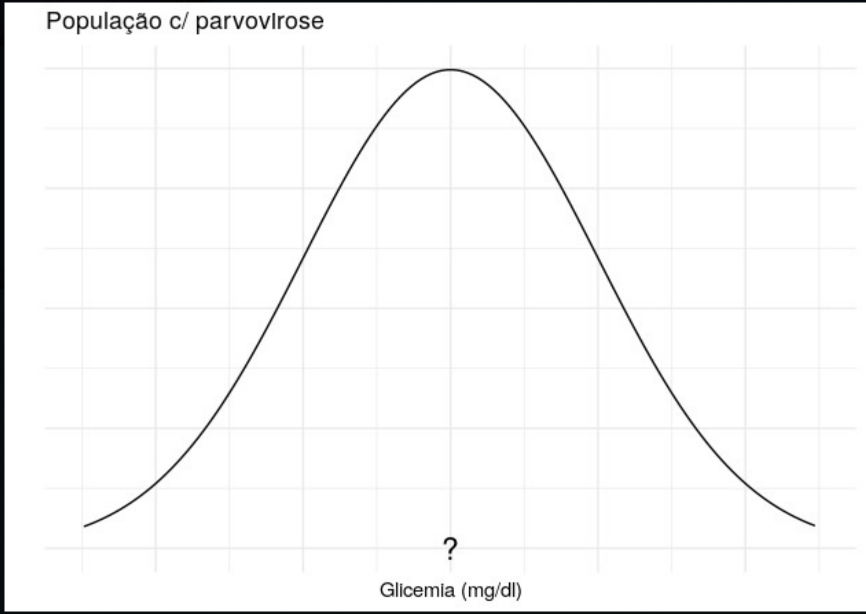
Amostra de cães com parvovirose (n = 30)
$\bar{x}$ = 104 mg/dl
s = 13 mg/dl

Exemplo 2

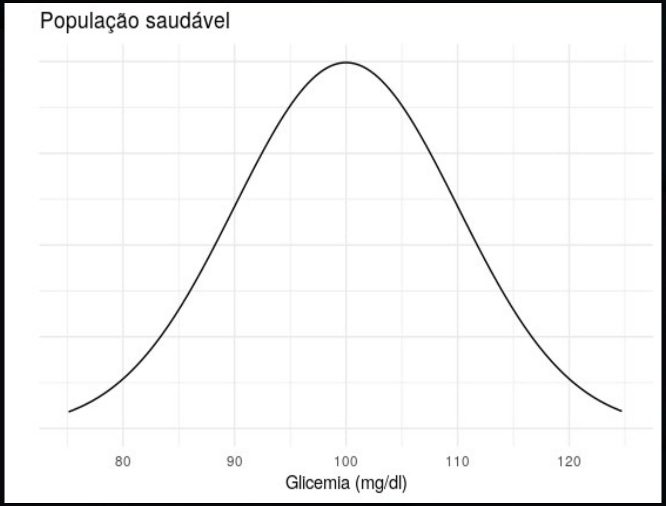
População saudável
μ = 100mg/dl
σ = 10 mg/dl



População com parvovirose
μ = ? mg/dl
σ = ? mg/dl

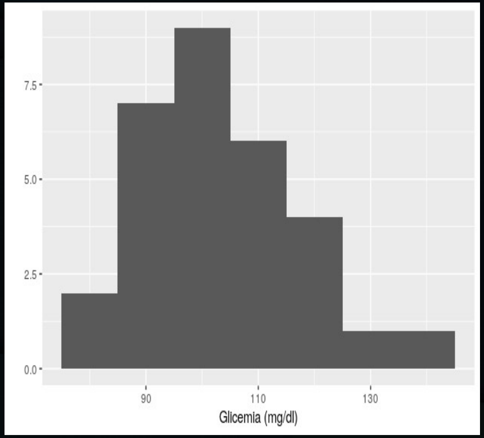


Exemplo 2



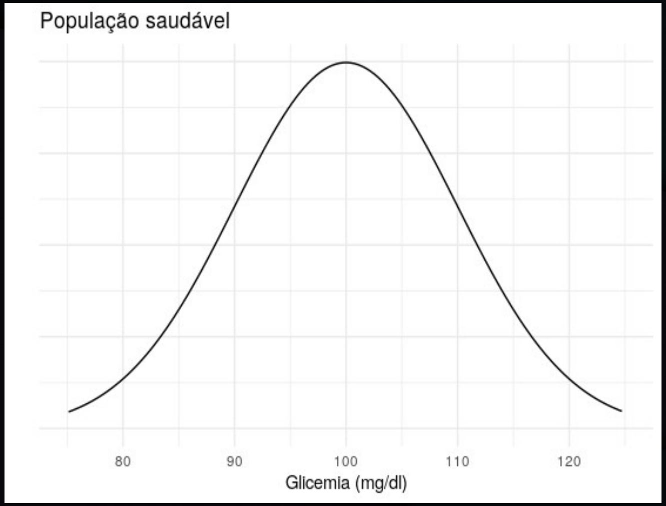
População
μ = 100mg/dl
σ = 10 mg/dl

É possível que uma amostra dessa População tenha $\bar{x}$ = 104 mg/dl ?

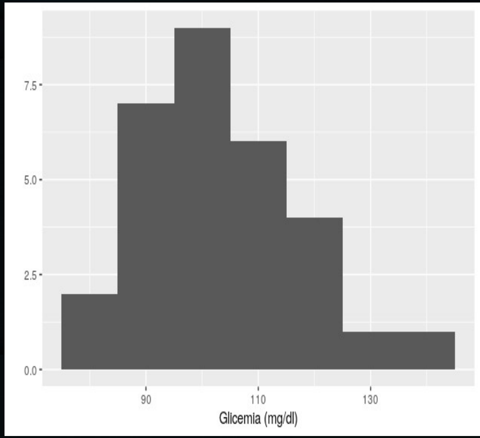


Amostra de cães com parvovirose (n = 30)
$\bar{x}$ = 104 mg/dl
s = 13 mg/dl

Exemplo 2



População
μ = 100mg/dl
σ = 10 mg/dl

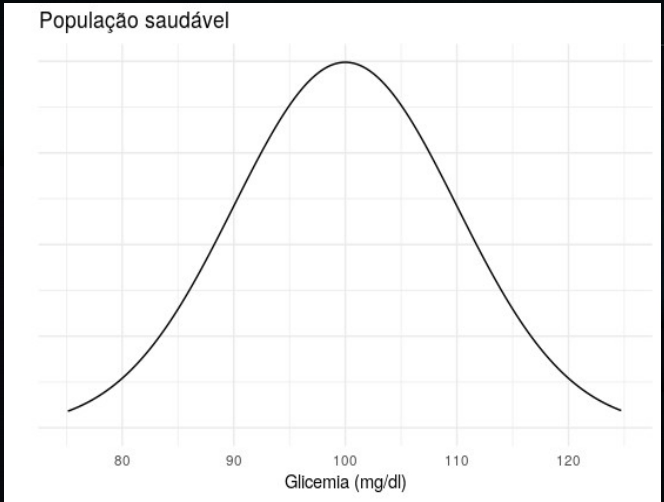


É possível que uma amostra dessa População tenha $\bar{x}$ = 104 mg/dl ?
Sim, parece bem razoável.
Quanto?

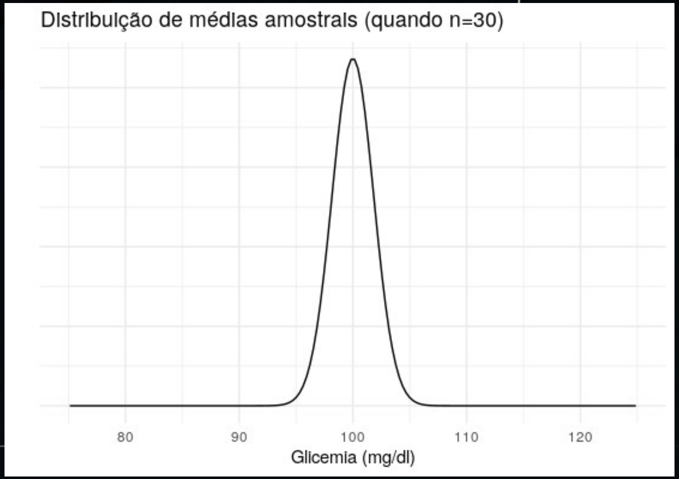
Amostra de cães com parvovirose (n = 30)
$\bar{x}$ = 104 mg/dl
s = 13 mg/dl

Exemplo 2

População
μ = 100mg/dl
σ = 10 mg/dl



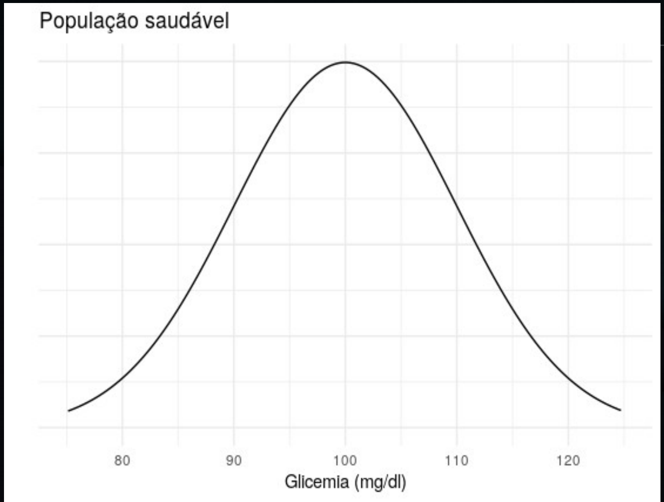
TLC

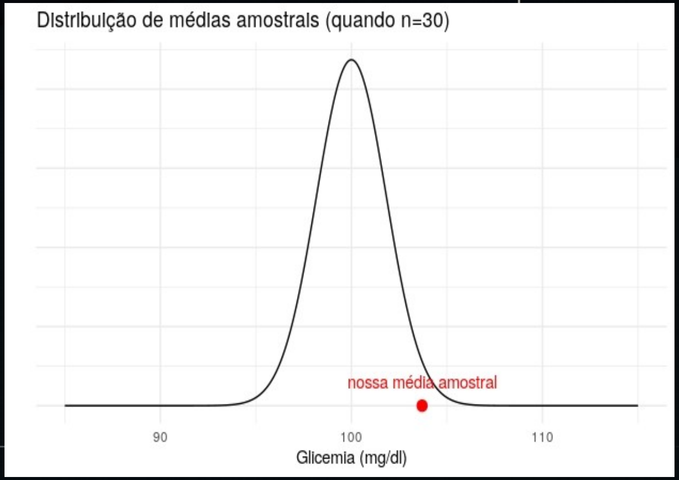


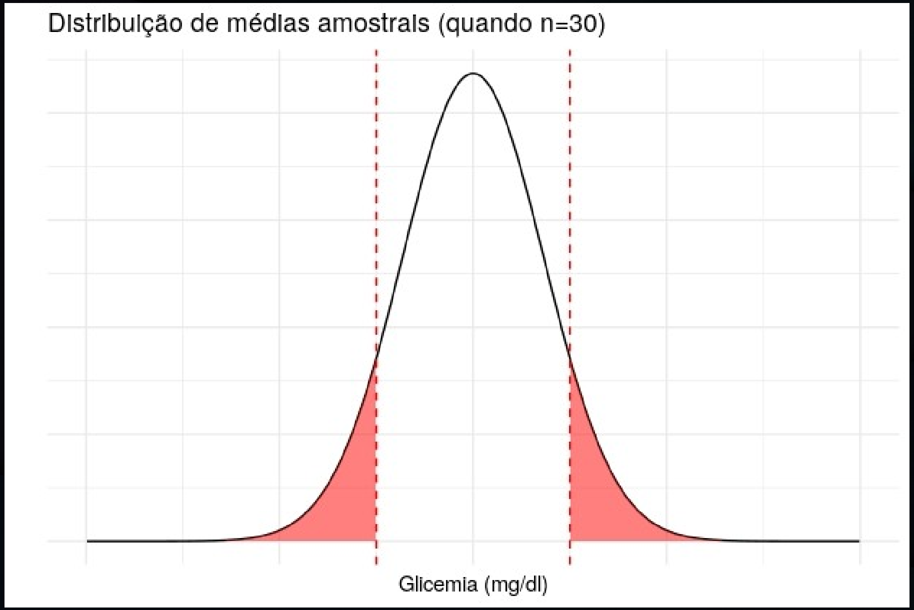
Exemplo 2

População
μ = 100mg/dl
σ = 10 mg/dl

TLC





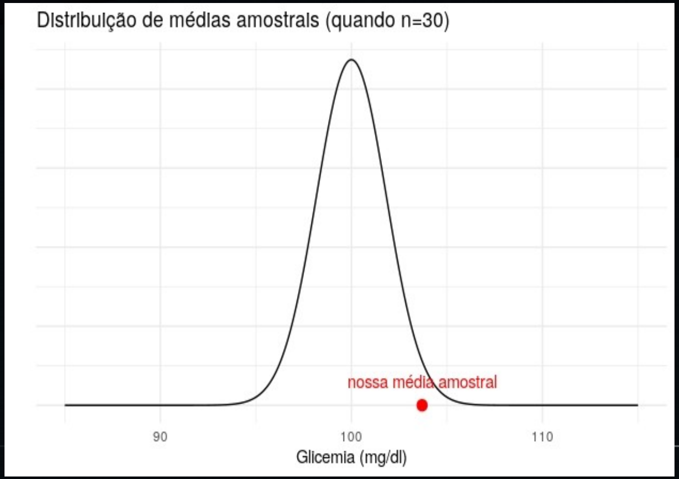


Nesse caso:
Probabilidade de uma $\bar{x}$ ser tão (ou mais) diferente de μ

Área embaixo da curva:
Probabilidade de uma média amostral ($\bar{x}$) estar nessa faixa de valores

Teoricamente possível, mas muito pouco provável.
Quanto? Posso usar o TLC e calcular pela normal.
Mas posso usar também a distribuição t de Student, que é mais fácil e não precisa de σ.

Resultado:
P-valor: 0,139 (ou 13,9%)



Formulação de hipóteses

H_0: μ = 100 mg/dl

H_1: μ ≠ 100 mg/dl

Procuramos evidência "suficiente" para rejeitar H_0.

Duas maneiras de rejeitar (ou não) H_0

Caso o IC **não inclua** o valor estipulado em H_0

P-valor < α

Rejeita H0

“Tenho evidência de que H_0 é falsa”

Caso o IC **inclua** o valor estipulado em H_0

P-valor >= α

Não rejeita H0

“Não tenho evidência de que H_0 é falsa”

Duas maneiras de rejeitar (ou não) H_0

H_0: $\mu = 100$ mg/dl

H_1: $\mu \neq 100$ mg/dl

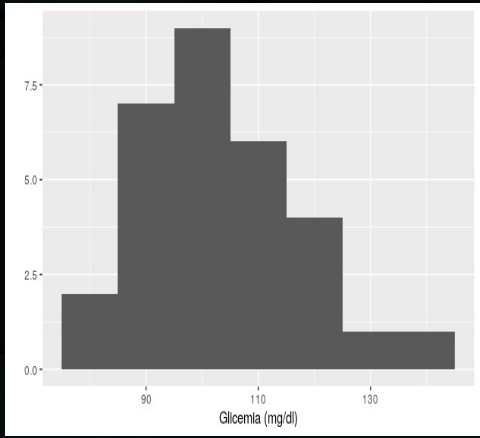
IC(μ,95%): 99 - 109

Inclui 100 mg/dl

P-valor: 0,139

P-valor > 5%

Não rejeita H_0



Amostra de cães com parvovirose (n = 30)
$\bar{x}$ = 104 mg/dl
s = 13 mg/dl

"Não tenho evidência de que H_0 é falsa"

Duas maneiras de rejeitar (ou não) H_0

H_0: $\mu = 100$ mg/dl

H_1: $\mu \neq 100$ mg/dl

IC(μ,95%): 99 - 109

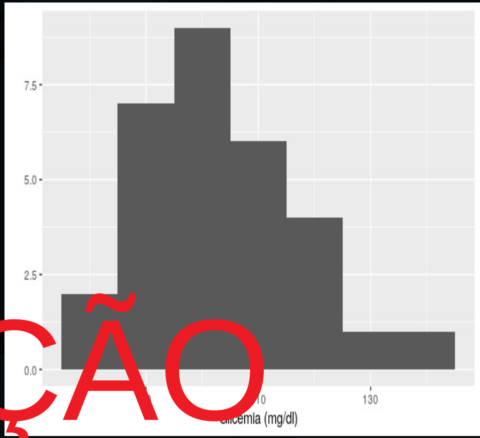
Inclui 100 mg/dl

P-valor: 0,139

P-valor > 5%

Não rejeita H_0

ATENÇÃO



Amostra de cães com parvovirose (n = 30)

$\bar{x}$ = 104 mg/dl

s = 13 mg/dl

"Não tenho evidência de que H_0 é falsa"

Conceitos

Hipóteses

- Hipótese Nula
  - H0
  - Hipótese em cima da qual o teste é construído
- Hipótese Alternativa
  - H1; Ha
  - Diferente da hipótese nula

Nível de significância (α)

É o complemento do nível de confiança

Assim como o nível de confiança, é escolhido pelo pesquisador, a depender do problema.

O mais comum é 5%. 1% e 0,1% também são utilizados.

Probabilidade de cometer um **Erro Tipo I**

Confiança	α
90%	10%
95%	**5%**
98%	2%
99%	1%
99,9%	0,1%

P-valor

- Também:
  - Valor-p; valor de p; p
  - Nível descritivo
  - Prob. de significância
- **Probabilidade** de obter o resultado observado, ou um mais extremo, **sob a hipótese de H_0** ser verdadeira

Lógica dos testes

Se p-valor < α

- Probabilidade de obter o resultado observado, sendo H_0 verdadeira, é **baixa**.
    - Duas explicações para ter acontecido:
        - H_0 é verdadeira, e você é um azarado!
        - H_0 é falsa.
- Conclusão: Rejeita H_0.

Lógica dos testes

Se p-valor > α

- Probabilidade de obter o resultado observado, sendo H_0 verdadeira, é **razoável**.
  - Duas explicações para ter acontecido:
    - H_0 é verdadeira, e tudo está funcionando normalmente.
    - H_0 é falsa, mas esse estudo não conseguiu evidência disso.
- Conclusão: Não rejeita H_0.

Erros Tipo I e II

	Realidade	
Resultado do teste	H0 verdadeira	H0 falsa
Rejeita H0 ($p < \alpha$)	Erro Tipo I	Acerto
Não rejeita H0 ($p > \alpha$)	Acerto	Erro Tipo II

Erro Tipo I: Rejeitar H0, quando ela é verdadeira.
Probabilidade de acontecer: α
Erro Tipo II: Não rejeitar H0, quando ela é falsa.
Probabilidade de acontecer: β

Poder do teste

Poder: 1 – β

Probabilidade de rejeitar H0, quando ela é falsa

Nos cenários onde H0 é falsa:

- Se β é a probabilidade de cometer o Erro Tipo II
  - (H0 é falsa mas não rejeitamos)
- então (1 – β) é a probabilidade de detectar que H0 é falsa (e rejeitar H0).

Só conseguimos calcular o Poder do Teste se “travarmos” uma H1, pois o cálculo do β depende do quanto H1 seria diferente de H0.

Testes de hipóteses

- Teste t
- ANOVA
- Regressão
- Mann-Whitney
- Wilcoxon
- Kruskal-Wallis
- Spearman
- Kolmorov-Smirnov
- Anderson-Darling
- Friedman
- …….. (e muitos, muitos, muitos outros)

Diferentes testes de hipóteses foram desenvolvidos para lidar com diferentes situações.

Tem:
teste pra média,
teste pra variância,
teste pra proporção,
teste pra diferença de médias,
teste pra saber se uma distribuição é normal, etc, etc, etc.

Como os diferentes testes de hipóteses foram construídos?

Construção de um teste de hipótese (Morettin, 2010)

Escolher um parâmetro qualquer θ, sobre o qual queremos fazer alguma afirmação (média, variância, proporção, diferença entre médias, etc).

Definir a hipótese nula e a alternativa. H0: $\theta = \theta_0$. H1 : $\theta \neq \theta_0$.
(ou H1: $\theta > \theta_0$; ou H1: $\theta < \theta_0$).

Usar a teoria estatística para calcular como o parâmetro escolhido θ (exemplo: $\theta = \mu$, ou $\theta = \mu_1 - \mu_2$) se distribui em amostras aleatórias (exemplo: $\bar{x}$ ou $\bar{x}_1 - \bar{x}_2$). Obter as propriedades dessa estatística (distribuição, média, desvio padrão).

Elaborar um procedimento (uma fórmula, por exemplo) que permita calcular o p-valor para uma amostra qualquer.

Falsificar hipóteses, não confirmá-las

- A lógica dos testes de hipóteses é contra intuitiva, pois eles não foram feitos para confirmar hipóteses
- Foram feitos para refutar (falsificar, rejeitar) hipóteses
- Hipóteses nula normalmente são aquelas em que os resultados podem ser atribuídos ao acaso

Escolha do teste estatístico apropriado

Escolha do teste

- Como escolher o teste de hipótese mais adequado para a pergunta a ser respondida?
- Diferentes testes de hipóteses foram desenvolvidos para lidar com diferentes situações.
- É necessário checar em quais situações cada teste é aplicável, e verificar se os dados atendem às premissas do teste escolhido.

Testes de hipóteses

- Teste t
- ANOVA
- Regressão
- Mann-Whitney
- Wilcoxon
- Kruskal-Wallis
- Spearman
- Kolmorov-Smirnov
- Anderson-Darling
- Friedman
- …….. (e muitos, muitos, muitos outros)

Diferentes testes de hipóteses foram desenvolvidos para lidar com diferentes situações.

Tem:
teste pra média,
teste pra variância,
teste pra proporção,
teste pra diferença de médias,
teste pra saber se uma distribuição é normal, etc, etc, etc.

Paramétricos x Não-Paramétricos

Os testes estatísticos podem ser divididos em:

- Testes **Paramétricos**:
  - Dados (ou população de estudo) precisam atender à certas premissas
  - Uma premissa comum, por exemplo, é que os dados tenham distribuição normal
- Testes **Não-Paramétricos**
  - Não fazem suposições sobre a distribuição e outros parâmetros da amostra ou população

- Sempre dê preferências aos testes Paramétricos: Quando as premissas desses são satisfeitas, eles possuem maior poder (menor erro Tipo II) do que os Não-Paramétricos.

Resumo
Testes de hipótese

- Descobrir um teste de hipótese que seja adequado ao seu problema.
- Verificar se você pode usar aquele teste, caso ele tenha premissas.
- Calcular a probabilidade da amostra analisada (ou uma mais extrema) ser obtida, dado que a hipótese nula é verdadeira (***p-valor***).
- Comparar essa probabilidade (***p-valor***) com o nível de significância (***α***).
- Caso a probabilidade seja baixa (***p-valor*** $< \alpha$) temos evidências de que a Hipótese Nula seja falsa.
- Caso a probabilidade seja alta (***p-valor*** $> \alpha$) não temos evidências de que a hipótese nula seja falsa
  - Atenção: Isso não quer dizer, necessariamente, que H0 é verdadeira.

Atenção!

- O que é **estatisticamente significante** pode não ser **biológica ou clinicamente significante** e vice-versa.
- *Ex*. Métodos de inseminação artificial (uma economia de 1 ou 2% pode ser uma diferença econômica grande, mas estatisticamente difícil de se obter)
- *Ex*. Dois diferentes anestésicos (pequenas variações na pressão sanguínea; a diferença pode ser estatisticamente significante, mas de pequena importância biológica)

Obrigado