II - Teste de hipótese para a média

Exemplo 1: Sabemos que a média do nível sérico de colesterol para a população de homens de 20 a 74 anos é 211 mg/100ml.

$\rightarrow$ O nível médio de colesterol da subpopulação de homens que são *fumantes hipertensos* é 211 mg/100ml?

Suponhamos que a distribuição dos níveis séricos de colesterol para a população de homens fumantes hipertensos é aproximadamente normal (média μ desconhecida e desvio padrão σ = 46 mg/100ml).

- Hipóteses estatísticas:

$$H: \mu = 211 \text{ mg/100ml}$$
$$A: \mu \neq 211 \text{mg/100ml}$$

Selecionamos uma a.a. de 12 homens desse grupo e seu nível sérico médio de colesterol é $\bar{x}$ = 217 mg/100ml. Essa média da amostra é compatível com a média suposta de 211 mg/100ml?

Sabemos que alguma quantidade de variabilidade amostral pode ser esperada.

O que acontece se a média da amostra é 230 mg/100ml ou 250 mg/100ml?

Quão longe de 211 precisa estar $\bar{x}$ antes que possamos concluir que μ seja realmente igual a algum outro valor?

Comparamos a média dessa amostra, $\bar{x}$, com a média 211; especificamente, queremos saber se a diferença entre a média da amostra e a média suposta é muito grande para ser atribuída somente ao acaso.

Se há evidência de que a amostra **não** pode vir de uma população com média 211, rejeitamos a hipótese nula.

Quando, dado que *H* é verdadeira, a probabilidade de se obter uma média da amostra tão extrema quanto o valor observado é *pequena*, rejeitamos *H*. Nesse caso, os dados não são compatíveis com a hipótese nula; eles dão mais suporte à hipótese alternativa *A*.

Em conseqüência concluímos que a média da população não pode ser 211. Dizemos que tal teste é ***estatisticamente significante***.

Note que a significância estatística não implica em significância clínica.

→ O que é uma probabilidade “pequena”?

Na maioria das aplicações, é escolhido 0,05.

Mais conservativos, escolhem uma probabilidade de 0,01.

Menos conservativos, uma probabilidade de 0,10 pode ser usada.

0,05 $\Rightarrow$ rejeitamos incorretamente 5% das vezes $\Rightarrow$ dados muitos testes repetidos de significância, 5 entre 100 erroneamente rejeitarão a hipótese nula quando ela é realmente verdadeira.

A probabilidade que escolhemos (0,05, 0,01 ou 0,10) é conhecida como ***nível de significância do teste*** de hipótese.

Resultado do teste	População	
	$\mu = \mu_0$	$\mu \neq \mu_0$
não rejeito *H*	correto	erro II
rejeito *H*	erro I	correto

Esses dois tipos de erros têm muito em comum com o resultado falso positivo e o resultado falso negativo que ocorrem nos testes de diagnóstico.

A probabilidade de se obter uma média tão ou mais extrema do que a média da amostra observada $\bar{x}$, dado que a hipótese nula é verdadeira, é chamada de **nível descritivo do teste (ou *P*-valor)** .

O *P*-valor é comparado ao nível de significância α predeterminado, para decidir se a hipótese nula deve ser rejeitada.

→ Como decidir?

Se $P \leq \alpha \Rightarrow$ rejeitamos *H.*
Se $P > \alpha \Rightarrow$ não rejeitamos *H.*

• Para conduzir um teste de hipótese usamos a distribuição amostral da média.

• Quando a população é normal com desvio padrão conhecido σ ou n é suficientemente grande, utilizamos um ***teste-z*** (Z é a *estatística do teste)*.

• Quando o desvio padrão da população não é conhecido, substituímos σ pelo valor da amostra s. E se a população original é normalmente distribuída temos um ***teste-t***.

Fixando nível de significância $\alpha = 0,05$.

$n = 12 \Rightarrow \bar{x} = 217$ mg/100ml $\Rightarrow P = 2 \times 0,326 = 0,625$

Conclusão: Como $P > \alpha$, não rejeitamos H

$\Rightarrow$ a evidência é insuficiente para concluir que o nível sérico médio de colesterol da população de fumantes hipertensos é diferente de 211 mg/100ml.

Há uma equivalência entre intervalos de confiança e testes de hipótese:

- A hipótese nula não será rejeitada quando μ_0 é qualquer valor que se encontra dentro do intervalo de confiança de 95% para μ.

- Embora os intervalos de confiança e os testes de hipóteses nos levem às mesmas conclusões, a informação fornecida por cada um é diferente:

  - o *intervalo de confiança* fornece uma série de valores razoáveis para o parâmetro μ e nos conta algo sobre a incerteza na nossa estimativa por ponto;

  - o *teste de hipótese* nos auxilia a decidir se o valor postulado da média é provável de estar correto ou não e fornece um *P*-valor específico.

Exemplo 2: Considere uma a.a. de 10 crianças selecionadas da população de bebês que recebem antiácidos que contêm alumínio.

A distribuição dos níveis de alumínio no plasma para essa população é aproximadamente normal com uma média μ e desvio padrão σ, desconhecidos.

n =10 $\Rightarrow \bar{x}$ = 37,20 μg/l (nível médio de alumínio no plasma para a amostra)

s = 7,13 μg/l (desvio padrão amostral)

O nível médio de alumínio no plasma para a população de bebês que não recebem antiácidos é 4,13 μg/l.

- Hipóteses: $H : \mu = 4{,}13$ μg/l
  $A : \mu \neq 4{,}13$ μg/l

Como não conhecemos o desvio padrão σ da população, utilizamos um teste-*t* ao invés de um teste-*z*.

nível de significância fixado $\alpha = 0{,}05$.

$P = 2 \times 0{,}0005 = 0{,}001 < \alpha \Rightarrow$ rejeitamos *H.*

Essa amostra de bebês fornece evidência de que o nível médio de alumínio no plasma das crianças que recebem antiácido *não é igual* ao nível médio de alumínio das crianças que não os recebem. O nível médio de alumínio é maior do que 4,13 μg/l.

Comentário:

Nos exemplos anteriores foram realizados testes bilaterais.

Algumas vezes temos interesse em testes unilaterais.

A decisão precisa ser tomada antes que uma amostra aleatória seja selecionada; ela não pode ser influenciada pelo resultado da amostra.

Exemplo 3: Considere a distribuição dos níveis de hemoglobina para a população de crianças com até 6 anos de idade que foram expostas a altos níveis de chumbo.

Essa distribuição tem uma média μ desconhecida; seu desvio padrão é assumido ser σ = 0,85 g/100ml.

Querermos saber se o nível médio de hemoglobina para essa população é igual à média da população geral de crianças com até 6 anos de idade μ = 12,29 g/100ml.

Acreditamos que se os níveis de hemoglobina das crianças expostas diferem daqueles das crianças não expostas, eles precisam ser menores do que a média ; em conseqüência, estamos preocupados somente com os desvios da média que estão abaixo de μ_0.

Desejamos testar as hipóteses,

$$H: \mu = 12{,}29$$

$$A: \mu < 12{,}29$$

(teste unilateral inferior ao nível de significância $\alpha = 0{,}05$)

Como σ é conhecido, usamos a distribuição normal ao invés da *t*.

Uma amostra aleatória de 74 crianças, que foram expostas a altos níveis de chumbo, apresentou um nível médio de hemoglobina $\overline{X} = 10{,}6$ g/100ml .

$P < \alpha \Rightarrow$ rejeitamos a hipótese nula.

Concluímos que o nível médio de hemoglobina para crianças que foram expostas ao chumbo é de fato mais baixo do que a média para crianças que não foram expostas.

Fixado $\alpha = 0,05 \Rightarrow$ obtemos a RC , β e poder.

A única maneira de diminuir α e β simultaneamente é reduzir a quantidade de sobreposição nas duas distribuições normais -- uma centrada em μ_0 e outra centrada em μ_1; quanto mais distantes estiverem os valores de μ_0 e de μ_1, maior o poder do teste.

Uma alternativa é aumentar o tamanho *n* da amostra. Aumentando-se *n*, diminuímos o erro padrão; isso faz com que as duas distribuições amostrais se tornem mais estreitas, reduzindo, portanto, a quantidade de sobreposição.
O erro padrão também diminui se diminuímos o desvio padrão σ da população original, mas isso usualmente não é possível.

Determinação do tamanho da amostra: Fixar α e β $\Rightarrow$ obter *n*

Diversos fatores influenciam a determinação do valor de *n*:

- se reduzimos a prob α do erro tipo I, isso resultará em um tamanho de amostra maior;
- se baixamos a prob β do erro tipo II, ou aumentamos o poder (1- β), isso produziria um valor maior de *n*;
- se consideramos uma média populacional da alternativa mais perto do valor da nula, a diferença $\mu_1 - \mu_0$ diminuirá e *n* aumentará; faz sentido necessitarmos de um tamanho de amostra maior para detectar uma diferença menor;
- quanto maior a variabilidade σ da população original, maior o tamanho da amostra exigido.

O tamanho da amostra para um teste bilateral é sempre maior do que o tamanho da amostra para o correspondente teste unilateral.