

Revisão

1. População e amostra
2. Tipos de variáveis e escala de medidas
3. Testes de hipóteses e p-valor

1. População e Amostra

<p>População</p> <p>Conjunto de todos os elementos (?) que contêm a(s) característica(s) de interesse.</p>	<p>Amostra</p> <p>Subconjunto da população.</p>
<p>Parâmetro</p> <p>Quantidade relacionada à população - usada para descrever a população. Exemplos</p>	<p>Estatística</p> <p>Qualquer função dos elementos da amostra. Exemplos</p>

A amostra deve ser:

- representativa (?)
- suficiente (?) e,
- aleatória (?)

2. Tipos de variáveis (características) e escala de medidas

qualitativas: nominal ou ordinal

quantitativas: discreta ou contínua

Nos métodos não-paramétricos o primeiro critério a ser verificado é a escala de mensuração.

Com esta podemos determinar a (“quantidade” de) informação contida nos dados e quais as maneiras de resumi-los - análises estatísticas apropriadas.



Escala de medidas

Níveis de mensuração:

- **nominal**: as respostas são nomes ou números (que não representam quantidades) => as operações aritméticas não fazem sentido
- **ordinal**: as respostas são nomes ou números (que não representam quantidades), como na escala nominal, porém podem ser ordenadas, segundo algum critério => as operações aritméticas não fazem sentido
- **intervalar**: as respostas são numéricas sem ponto zero natural (o zero é arbitrário) e os intervalos (diferenças) são significativos, mas as razões não são (temperatura de 0°C não significa falta de temperatura, e 20°C não significa o dobro de calor do que 10°C)
- **razão**: as respostas têm as propriedades da escala intervalar, e possuem um zero verdadeiro (representa uma origem absoluta e não há valor numérico negativo – 0 significa “falta de”); tanto os intervalos como as razões são significativos => todas as operações aritméticas podem ser aplicadas.

Por fim,

- uma variável pode ser especificada de formas diferentes => importância da escala: como queremos obter as informações (quanto de informação queremos extrair no estudo!)
- a escala de razão é a de nível mais alto:

(mais informativa) **razão => intervalar => ordinal => nominal** (menos informativa)

3. Testes de hipóteses e p-valor: contexto de inferência estatística

- Hipótese estatística: declaração a respeito de um ou mais parâmetros. Esta é testada com evidências contidas na amostra:
 - se rejeitada => a evidência obtida na amostra não é suficiente para afirmarmos, com certo grau de confiança, que a hipótese é verdadeira
 - se aceita => a evidência obtida na amostra é suficiente para afirmarmos, com certo grau de confiança, que a hipótese é verdadeira.
- Teste de hipótese (t.h.): para responder uma questão estatística, tal questão é **transformada** em uma hipótese, usando parâmetros – declaração que podemos testar, via teste estatístico, cujo resultado nos leva a aceitar, ou não, a hipótese.
- Teste estatístico: uma função da amostra e do parâmetro (em questão) usado para a tomada de decisão.

(observe que no teste fazemos inferência, isto é, usamos a estatística para inferir a respeito do parâmetro e, aceitamos ou não a declaração a respeito da população)

- Como realizar um t.h.?

a) construir as hipóteses

- H_0 : hipótese a ser testada

(se desejamos determinar se a declaração a respeito da população é falsa, fazemos desta a hipótese nula)

- H_1 : hipótese alternativa: complementar à H_0

(se desejamos determinar se a declaração a respeito da população é verdadeira, fazemos desta a hipótese alternativa)

b) seleccionar o teste estatístico: relacionado à quantidade populacional.

c) aplicar o teste e tomar a decisão: rejeitar ou não H_0 .

d) Pontos importantes:

- região crítica (r.c.): valores no espaço amostral que nos levam a rejeitar H_0 .

- nível de significância: $\alpha = P[\text{rej. } H_0 \mid H_0 \text{ verd.}]$

- nível crítico: **p-valor**: menor nível de significância no qual H_0 seria rejeitada (medida intuitiva da força dos dados contra H_0).

Observações:

- As hipóteses são especificadas em termos de parâmetros e a escolha do teste depende do parâmetro a ser testado e de como os valores amostrais foram registrados (natureza da variável).
- Usando o teste estatístico e definindo a região crítica (ou calculando o p-valor), tomamos uma decisão => não existe evidência suficiente na amostra para rejeitar/aceitar H_0 .
- Não rejeitar a hipótese nula não significa dizer (ou afirmar) que a hipótese alternativa é válida com o mesmo nível de confiança.

Quais os tipos de hipóteses?

Alguns tipos de hipóteses:

Denotando por θ o parâmetro de interesse (média, variância, proporção, mediana, *etc.*) e θ_0 um valor especificado de θ :

a) $H_0: \theta = \theta_0$ *versus* $H_1: \theta \neq \theta_0$

b) $H_0: \theta \geq \theta_0$ *versus* $H_1: \theta < \theta_0$

c) $H_0: \theta \leq \theta_0$ *versus* $H_1: \theta > \theta_0$

(observe que o sinal “=” está sempre em H_0).

Exemplos:

1. Produção de peças: uma amostra foi selecionada a fim de verificar se a máquina, que produz as peças, está sob especificação, isto é, se esta produz, no máximo, 5% de itens que apresentam algum tipo de defeito.

H_0 : a máquina está em condições de se operada

H_1 : a máquina não está em condições de se operada

$\Rightarrow H_0: p \leq 0,05 \text{ versus } H_1: p > 0,05.$

2. A fim de verificar se existe diferença de comportamento entre crianças que passaram pela educação infantil (até cinco anos) e crianças que entraram (direto) na educação básica (a partir de cinco anos), uma escola aplicou um teste nas crianças.

H_0 : não existe diferença comportamental

H_1 : existe diferença comportamental.