Introdução à Bioestatística

Marcelo Goulart Correia

Instituto Nacional de Cardiologia

February 26, 2016

- Situações que podem existir uma teoria preconcebida sobre a população submetida ao estudo
- Investigação de duas hipóteses (Nula e Alternativa) que retratam esse conceito pré-estabelecido
- Ou seja:
    - Supondo a priori que a distribuição da população seja conhecida
    - Extrai-se uma amostra aleatória da população alvo
    - **Se a distribuição dessa amostra difere da distribuição preconcebida da população, então a suposição inicial é rejeitada**

- Exemplo
  - Suponha um estudo onde queremos comparar a altura dos moradores do Nordeste com o restante do Brasil. Antes de coletar os dados, desenha-se a suposição a priori
    - H_0: A altura média não difere do restante do país (1,73)
    - H_A: A altura média difere do restante do país (1,73)
  - Coletam-se duas amostras de tamanhos iguais da população do Nordeste
    - 1: {1,50; 1,52; 1,48; 1,55; 1,60; 1,49; 1,55; 1,63} = 1,54
    - 2: {1,65; 1,80; 1,73; 1,52; 1,75; 1,65; 1,75; 1,78} = 1,70
  - Caso 1 –> Se a amostra for representativa, a hipótese nula deve ser rejeitada
  - Caso 2 –> Dado o tamanho da amostra, não se tem segurança para rejeitar ou não rejeitar a hipótese nula

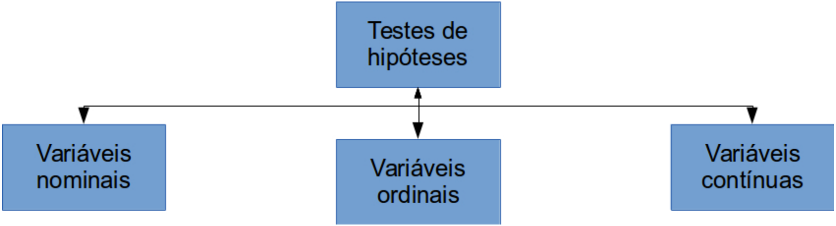
- Num teste de hipótese:
  - Decidir se uma hipótese nula (H_0) pode ou não ser rejeitada
  - É necessário estabelecer uma hipótese alternativa (H_A) que será admitida quando H_0 for rejeitada
  - Definir a estatística relacionada com a hipótese que será testada (estatística teste)
  - Escolher o nível de significância do teste
  - Se H_0 for correto, o teste de hipótese resultará numa resposta errada com probabilidade α
  - α = probabilidade de uma amostra ter valores atípicos

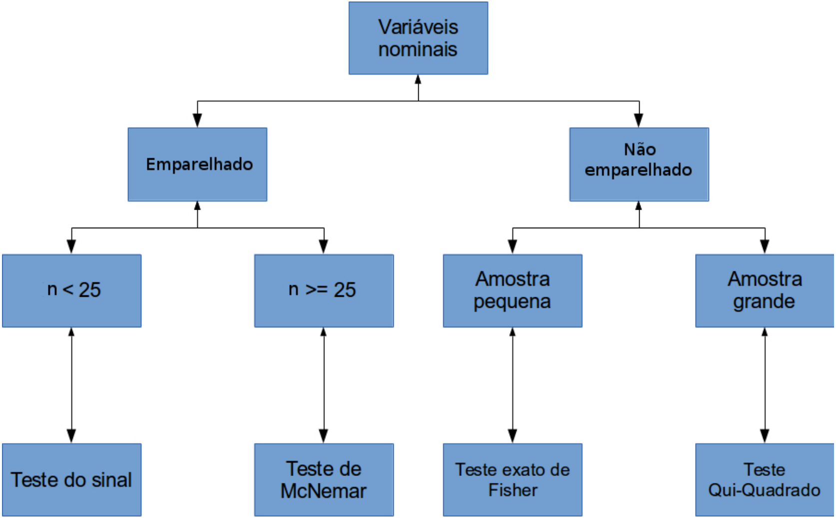
- Cuidados:
  - Conclusão sobre uma hipótese pode ser incorreta quando:
    - Erro Tipo I: Rejeitar a hipótese nula quando a mesma é correta
    - Erro Tipo II: Não rejeitar a hipótese nula quando a mesma é incorreta
  - Quanto menor for α maior será o β –> Explosão do tamanho amostral
  - Simplifique a hipótese nula
  - Escolha a estatística teste correta
  - Cuidado com a consequência de seu julgamento sobre a hipótese nula

- Exemplo
  - Teste do efeito de um tratamento experimental:
  - H_0: Paciente piora ou permanece estável no tratamento
  - H_1: Paciente melhora no tratamento

  - Teste de segurança de uma nova máquina de hemodiálise:
  - H_0: Máquina menos eficiente ou equivalente
  - H_1: Máquina mais eficiente

- Os testes estatísticos podem ser aplicados para dados nominais, ordinais e numéricos
- Alguns testes permitem o cruzamento de variáveis com características distintas (Nominal x Numérica, Ordinal x Numérica)
- Um teste pode ser unicaudal ($<$, $\leq$, $>$, $\geq$) ou bicaudal ($=$, $\neq$)
  - **CUIDADO**: "Se um teste de não inferioridade é significativo, logo apresenta superioridade"
- Os testes estatísticos para dados numéricos dividem-se em dois grandes grupos:
  - Testes paramétricos –> Dependem de pressupostos para serem utilizados
  - Testes não-paramétricos –> Não depende de pressupostos para serem utilizados


Testes de hipóteses
Variáveis nominais
Variáveis ordinais
Variáveis contínuas


Variáveis nominais
Emparelhado
Não emparelhado
n < 25
n >= 25
Amostra pequena
Amostra grande
Teste do sinal
Teste de McNemar
Teste exato de Fisher
Teste Qui-Quadrado

- Emparelhamento
  - Natural –> Comparação de irmãos gêmeos
  - Artificial –> Comparação de indivíduos com características similares
  - Auto-emparelhamento –> Um elemento atua como seu próprio controle
- **Atenção:** Não é porque dois grupos tem o mesmo tamanho amostral (n) que os dados serão emparelhados

- Teste do sinal
  - Recebe esse nome pois a diferença de cada par selecionado é convertido para sinais (+, - ou 0) não importando a magnitude da diferença
  - H_0: Encontrar a mesma quantidade de sinais positivos e negativos
  - H_A: Não encontrar a mesma quantidade de sinais positivos e negativos

$$\boxed{Z = 2(p' - 0,5)\sqrt{n}} \quad (1)$$

- p' = Proporção de + ou -

- Exemplo
  - Deseja-se escolher dentre dois equipamentos (A e B) que realizam a mesma tarefa. São realizadas 12 tarefas diferentes na máquina A e as mesma 12 na máquina B. O objetivo é verificar qual máquina realiza mais vezes a mesma tarefa num espaço de tempo qualquer.

Tarefa	Máquina A	Máquina B	Diferença (A-B)
1	40	29	+
2	22	16	+
3	22	29	-
4	45	41	+
5	68	61	+
6	33	24	+
7	48	54	-
8	75	68	+
9	41	36	+
10	44	36	+
11	47	42	+
12	31	25	+

- $p' = 2/12 = 1/6$
- $Z = 2 * (1/6 - 1/2) * \sqrt{12} = -2,28$
- Valor crítico de Z ($\alpha = 0,05$) = $\pm$ 1,96 (Bicaudal)
- Existe diferença entre os equipamentos para um nível de significância de 5%

- Exercício

- Teste de McNemar
  - É uma extensão do teste Qui-Quadrado para amostras emparelhadas
  - H_0: As frequências encontradas são iguais
  - H_A: As frequências encontradas são diferentes

$$\boxed{Q^2_{obs} = \frac{(a-d)^2}{(a+d)}} \tag{2}$$

- a = Frequência de casos que evoluiram
- d = Frequência de casos que regrediram

- Exemplo
  - 1319 pacientes foram interrogados sobre a prevalência de resfriados aos 12 anos de idade e posteriormente aos 14 anos de idade. Aos 12 anos de idade, o percentual de atingidos pela enfermidade foi de 27% e aos 14 anos de idade 35.5%. Houve aumento significativo da prevalência de resfriados?

Resf. 12 anos \Resf. 14 anos	Sim	Não	Total
Sim	212	144	356
Não	256	707	963
Total	468	851	1319

$$Q^2_{obs} = \frac{(144-256)^2}{(144+256)} = 12544/400 = 31,36$$

- Valor crítico de Q^2 ($\alpha = 0,05$ e GL $= 1$) $=$ 3,84
- Existe diferença entre as frequências de resfriado para um nível de significância de 5%

- Exercício

- Teste de Qui-Quadrado
  - Utilizado para comparar duas variáveis nominais independentes (não emparelhadas)
  - H_0: Não existe diferença entre as variáveis analisadas
  - H_A: Existem diferenças para as variáveis analisadas

$$\boxed{\chi^2 = \Sigma \frac{(\text{observado - esperado})^2}{\text{esperado}}} \tag{3}$$

- Esperado $= \frac{\text{Total de um desfecho}}{\text{Total amostral}} * \text{Total de uma resposta}$

- Exemplo
  - Um estudo que deseja-se analisar a mortalidade de uma determinada doença após o surgimento de um novo tratamento

Tratamento \Óbito	Sim	Não	Total
Tratamento clássico	35	24	59
Tratamento novo	70	21	91
Total	105	45	151

- Valor esperado para sobrevida

$$\text{Tratamento clássico} = (105/150) * 59 = 41,3$$
$$\text{Tratamento novo} = (105/150) * 91 = 63,7$$

- Valor esperado para óbito

$$\text{Tratamento clássico} = (45/150) * 59 = 17,7$$
$$\text{Tratamento novo} = (45/150) * 91 = 27,3$$

$$\chi^2 = \frac{(35-41{,}3)^2}{41{,}3} + \frac{(24-17{,}7)^2}{17{,}7} + \frac{(70-63{,}7)^2}{63{,}7} + \frac{(21-27{,}3)^2}{27{,}3} = 5,2803$$

- Valor crítico de Q ($\alpha = 0,05$) = 3,84
- Existe diferença entre os tratamentos para um nível de significância de 5%

- Exercício

- Teste Exato de Fisher
  - Utilizado para comparar duas variáveis nominais independentes (não emparelhadas)
  - Deve ser utilizado quando 25% das respostas possíveis tiver frequência menores ou iguais a 5
  - H_0: Não existe diferença entre as variáveis analisadas
  - H_A: Existem diferenças para as variáveis analisadas
- O método consiste em:
  - Criar matrizes para cada decréscimo de valor da menor célula da tabela (Afastamento da hipótese nula)
  - Calcular o resultado de cada matriz

$$\boxed{P_a = P[X = a] = \frac{(a+b)! + (c+d)! + (a+c)! + (b+c)!}{a!b!c!d!n!}} \tag{4}$$

  - Somar os resultados e verificar o valor nível de significância

- Exemplo
  - Doentes psiquiátricos podem ser classificados em psicóticos e neuróticos
  - Realiza-se um estudo sobre os sintomas suicidas em duas amostras de 20 doentes de cada grupo

Sintoma suicida \Tipo de doente	Psicótico	Neurótico	Total
Presente	2	6	8
Ausente	18	14	32
Total	20	20	40

$$\begin{pmatrix} 2 & 6 \\ 18 & 14 \end{pmatrix} \begin{pmatrix} 1 & 7 \\ 19 & 13 \end{pmatrix} \begin{pmatrix} 0 & 8 \\ 20 & 12 \end{pmatrix}$$

$$P_a = P_2 + P_1 + P_0 = 0,1176$$

- Não existe diferença entre os tipos de doenças para um nível de significância de 5%

- Exercício