Introdução à Bioestatística

Marcelo Goulart Correia

Instituto Nacional de Cardiologia

May 11, 2015

1. Testes de hipóteses para dados numéricos

- É importante verificar qual a distribuição dos dados
- Os testes paramétricos precisam atender ao requisito de normalidade

- Teste de Shapiro Wilk
  - Surge em 1965
  - É baseado na estatística W
  - H_0: A amostra provem de uma distribuição normal
  - H_A: A amostra não provem de uma distribuição normal

$$\boxed{W = \frac{b^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}} \tag{1}$$

$$\boxed{b = \begin{cases} \sum_{i=1}^{n} /2a_{n-i+1} * (x_{n-i+1} - x_i), \textbf{ se n é par} \\ \sum_{i=1}^{n} +1/2a_{n-i+1} * (x_{n-i+1} - x_i), \textbf{ se n é impar} \end{cases}} \tag{2}$$

- Teste de Kolmogorov-Smirnov
  - Alternativa ao teste de Shapiro-Wilk
  - É baseado na função distribuição acumulada

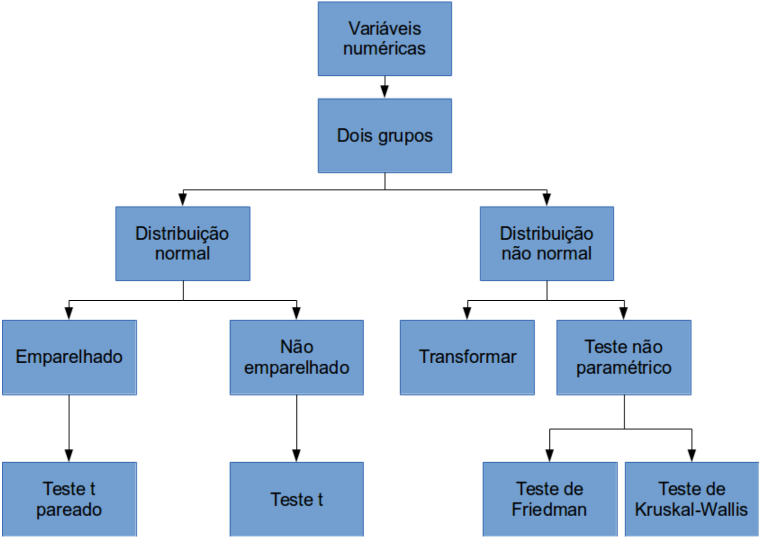
  - H_0: A amostra provém de uma distribuição normal
  - H_A: A amostra não provém de uma distribuição normal

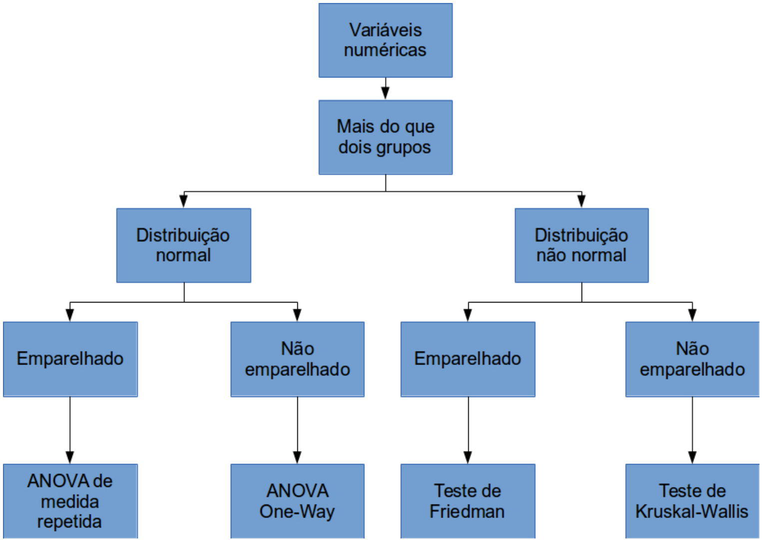
$$\boxed{D_n = max(D_n^+, D_n^-)} \tag{3}$$

$$\boxed{D_n^+ = \mathbf{max}(F_n(x) - F(x))} \tag{4}$$

$$\boxed{D_n^- = \mathbf{max}(F(x) - F_n(x))} \tag{5}$$

- Cuidados
  - Alguns testes tem limitação de tamanho amostral
  - Com amostras grandes qualquer desvio resulta numa rejeição da hipótese nula
  - Verifique outros aspectos da variável (média, mediana, IQR, intervalo de confiança, histograma)


Variáveis numéricas
Dois grupos
Distribuição normal
Distribuição não normal
Emparelhado
Não emparelhado
Transformar
Teste não paramétrico
Teste t pareado
Teste t
Teste de Friedman
Teste de Kruskal-Wallis


Variáveis numéricas
Mais do que dois grupos
Distribuição normal
Distribuição não normal
Emparelhado
Não emparelhado
Emparelhado
Não emparelhado
ANOVA de medida repetida
ANOVA One-Way
Teste de Friedman
Teste de Kruskal-Wallis

- Teste do t emparelhado
  - Utilizada para a comparação de médias entre dois grupos emparelhados
  - H_0: Não existem diferenças entre as médias observadas nos dois grupos
  - H_A: Existem diferenças entre as médias observadas nos dois grupos

$$\boxed{T = \frac{\bar{D} - \mu_D}{s_D/\sqrt{n}}} \tag{6}$$

$$\boxed{s_D^2 = \frac{\sum_{i=1}^{n}(D_i - \bar{D})^2}{n-1}} \tag{7}$$

- Exemplo
  - Para determinar a eficiência de um antitérmico, foram mensuradas as temperaturas de 20 indivíduos antes e após o consumo do medicamento

Indivíduo	Antes	Depois	Indivíduo	Antes	Depois
1	37,5	37,8	11	39,3	38
2	36	36,3	12	37,5	37,1
3	39	37,6	13	38,5	36,6
4	38	37,2	14	39	37,5
5	37,8	36,9	15	36,9	37
6	38,5	37,7	16	37	36,2
7	36,9	36,8	17	38,5	37,6
8	39,4	38,1	18	39	36,8
9	37,2	36,7	19	36,2	36,4
10	38,1	37,3	20	36,8	36,8

- Exemplo

Indivíduo	Diferença	Indivíduo	Diferença
1	-0,3	11	1,3
2	-0,4	12	0,4
3	1,4	13	1,9
4	0,8	14	1,5
5	0,9	15	-0,1
6	0,8	16	0,8
7	0,1	17	0,9
8	1,3	18	2,2
9	0,5	19	-0,2
10	0,8	20	0

- Exemplo
  - $\bar{D} = 0,73$ // $s_D = 0,7356$
  - $T_{obs} = \frac{0,73-0}{0,7356/\sqrt{20}} = 4,438$
  - Valor crítico de T ($\alpha = 0,05$) = 2,093 (Bicaudal)
  - Existe diferença entre os pares para um nível de significância de 5%

- Exercício

- Teste do t
  - Utilizada para a comparação de médias entre dois grupos independentes
  - H_0: Não existem diferenças entre as médias observadas nos dois grupos
  - H_A: Existem diferenças entre as médias observadas nos dois grupos
- O método consiste em:
  - Calcular a média dos dois grupos ($\bar{x}_1$ e $\bar{x}_2$)
  - Calcular a variância dos dois grupos (s_1^2 e s_2^2)
  - Calcular a variância ponderada

$$\boxed{s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} \tag{8}$$

  - Calcular o valor de t

$$\boxed{t = \frac{\bar{x}_2 - \bar{x}_1}{\sqrt{s^2(1/n_1 + 1/n_2)}}} \tag{9}$$

- Exemplo
  - Como forma de verificar se duas dietas são igualmente eficientes, foram criados dois grupos de paciente para duas dietas distintas. As perdas de peso foram coletadas conforme descrito abaixo

Dieta 1	Dieta 2
12	15
8	19
15	15
13	12
10	13
12	16
14	15
11	
12	
13	

- Exemplo

$$\bar{x}_1 = 12 + 8 + ... + 13/10 = 12 \ // \ \bar{x}_2 = 15 + 19 + ... + 15/7 = 15$$
$$s_1^2 = (1476 - 120^2/10)/9 = 4 \ // \ s_2^2 = (1605 - 105^2/7)/6 = 5$$
$$s^2 = 9 * 4 + 6 * 5/9 + 6 = 4,4$$
$$t = 15 - 12/\sqrt{4,4 * (1/10 + 1/7)} = 2,902$$

- Valor crítico de T ($\alpha = 0,05$) = 2,131 (Bicaudal)
- Existe diferença entre as dietas para um nível de significância de 5%

- ANOVA One-Way (Um fator)
  - Utilizada para a comparação de médias entre mais de dois grupos
  - H_0: Não existem diferenças entre as médias observadas nos grupos
  - H_A: Existem diferenças entre as médias observadas nos grupos

Fonte de Variação	Graus de liberdade	Soma dos quadrados	Quase variâncias	F0
Entre tratamentos	t-1	SQE	QVE = SQE/(t-1)	QVE/QVD
Dentro dos tratamentos	N-t	SQD = SQT – SQE	QVD = SQD/(N-t)	
Total	N-1	SQT		

- Exemplo
  - Três tratamentos são utilizados para uma determinada enfermidade, deseja-se verificar se existem diferenças entre os resultados obtidos em cada tratamento

A	B	C
15	23	19
10	16	15
13	19	21
18	18	14
15		16
13		
—	—	—
84	76	85

- Exemplo
  - Graus de liberdade

$$\text{Total} \to 15 - 1 = 14$$
$$\text{Tratamentos} \to 3 - 1 = 2$$
$$\text{Resíduos} \to 14 - 2 = 12$$

  - Média do quadrado da soma dos valores

$$C = (84 + 76 + 85)^{/}15 = 4001,67$$

  - Soma dos quadrados

$$SQT = 15^2 + 10^2 + ... + 16^2 - 4001,67 = 159,33$$
$$SQE = \frac{84^2}{6} + \frac{76^2}{4} + \frac{85^2}{5} - 4001,67 = 63,33$$
$$SQD = 159,33 - 63,33 = 96$$

  - Quase variâncias

$$QVE = 63,33/2 = 31,67$$
$$QVD = 96/12 = 8$$

  - Estatística F

$$F = 31,37/8 = 3,96$$

- Exemplo
  - Valor crítico de F ($\alpha = 0,05$, $GLE = 2$, $GLD = 12$) = 3,89
  - Existe diferença entre os tratamentos para um nível de significância de 5%

- Exercício

- ANOVA One-Way (Um fator) de medida repetida
  - Utilizada para a comparação de médias entre mais de dois grupos pareados
  - H_0: Não existem diferenças entre as médias observadas nos grupos
  - H_A: Existem diferenças entre as médias observadas nos grupos
- A tabela conterá:
  - A interação entre variável e grupos
  - Erros de medidas dessa interação

- Exemplo
  - É obtida a medida de proteína c-reativa em três momentos (Pré-treinamento, duas semanas de treinamento, fim do treinamento) num grupo de pacientes.

Descriptive Statistics

	Mean	Std. Deviation	N
Pre_Training	3.0875	.97763	20
Week2	2.9675	.89388	20
Post_Training	2.2450	.49892	20

- Exemplo

Tests of Within-Subjects Effects

Measure:CRP

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
time	Sphericity Assumed	8.308	2	4.154	21.032	.000	.525
	Greenhouse-Geisser	8.308	1.171	7.092	21.032	.000	.525
	Huynh-Feldt	8.308	1.202	6.912	21.032	.000	.525
	Lower-bound	8.308	1.000	8.308	21.032	.000	.525
Error(time)	Sphericity Assumed	7.505	38	.198			
	Greenhouse-Geisser	7.505	22.257	.337			
	Huynh-Feldt	7.505	22.837	.329			
	Lower-bound	7.505	19.000	.395			

- Exercício