USP – ICMC – SME0810 - Métodos Não Paramétricos
1ª prova – 2º/2014 – 7/10/2014

S o l u ç ã o

1. Um experimento foi realizado com o objetivo de avaliar o efeito da intoxicação por álcool sobre a coordenação. Nove pessoas ingeriram uma certa quantidade de álcool e após certo tempo escreveram uma frase tantas vezes quanto foram capazes em um intervalo de um minuto. O número de palavras corretas foi contado e transformado em um escore de modo que o valor 0 representa o resultado de uma pessoa sem influência de álcool. Valores positivos e negativos representam acréscimo e decréscimo, respectivamente, em relação ao resultado de uma pessoa em condições normais. Os escores foram 10, -8, -6, -2, 15, 0, -7, 5, -8. (a) Apresente uma estimativa para o escore mediano. (b) Apresente um intervalo de confiança com coeficiente próximo a 95% para o escore mediano. (c) Apresente a probabilidade de cobertura do intervalo de confiança do item (b). (d) Pode ser afirmado que há efeito da intoxicação por álcool sobre a coordenação das pessoas?
Solução. Os escores x em ordem crescente são
-8, -8, -7, -6, -2, 0, 5, 10 e 15.
(a) Supomos que a distribuição do escore é contínua com mediana única. A mediana amostral é um estimador do escore mediano. Como $n = 9$, $X_{(5)}$ corresponde à mediana amostral. A estimativa é $x_{(5)} = -2$.
(b) Um intervalo de confiança (IC) para o escore mediano pode ser obtido com base na estatística de teste do sinal. Para um coeficiente de confiança de $1 - \alpha = 0,95$, inicialmente calculamos o quantil superior $1 - \alpha/2 = 0,975$ da distribuição binomial($n = 9, 1/2$), denotado por $b_{1-\alpha/2}$. A função massa de probabibilidade e a função distribuição acumulada são dadas abaixo.

j	0	1	2	3	4	5	6	7	8	9
f(j)	0.001953	0.01758	0.07031	0.1641	0.2461	0.2461	0.1641	0.07031	0.01758	0.001953
F(j)	0.001953	0.01953	0.08984	0.2539	0.5000	0.7461	0.9102	0.98047	0.99805	1.000000

Consultando a linha da função distribuição acumulada acima obtemos $b_{1-\alpha/2} = 7$. A posição do limite superior do IC é dada por $b_{1-\alpha/2} = 7$. A posição do limite inferior do IC é dada por $C_\alpha = n + 1 - b_{1-\alpha/2} = 9 + 1 - 7 = 3$. O intervalo de confiança é dado por (X_3, X_7). Com base nos dados calculamos $\left(x_{(3)}, x_{(7)}\right) = (-7, 5)$.
(c) Utilizamos a função massa de probabibilidade dada acima. A probabilidade de cobertura do intervalo é $\sum_{j=3}^{7} \mathrm{P}(X = j) = 0,891$.
(d) Com base no intervalo de confiança do item (c), que inclui o valor 0, não pode ser afirmado que há efeito da intoxicação por álcool sobre a coordenação das pessoas.

2. Nove laboratórios participaram de um estudo com o objetivo de verificar se a dose efetiva mediana de uma certa droga ultrapassa 0,5. Os dados coletados foram 0,41, 0,52, 0,91, 0,45, 1,06, 0,82, 0,78, 0,68 e 0,75. (a) Apresente um teste com nível de siginificância nominal de 5%. O teste é conservador ou liberal? (b) Apresente uma solução com base no valor-p do teste.
Solução. A um nível de significância nominal $\alpha = 0,05$, deve ser tetada $\mathrm{H}_0 : \theta = \theta_0$ contra $\mathrm{H}_1 : \theta > \theta_0$, em que θ denota a dose efetiva mediana e $\theta_0 = 0,5$. Não há empates e nenhuma observação é igual ao valor de teste (θ_0). Supomos que a distribuição da dose efetiva é contínua com mediana única.
(a) A solução está baseada no teste do sinal. Inicialmente calculamos o quantil $1-\alpha = 0,95$ da distribuição binomial($n = 9, 1/2$), denotado por $b_{1-\alpha}$. Consultando a função distribuição acumulada apresentada na questão 1, obtemos $b_{1-\alpha} = 7$. Consultando a função massa de probabilidade apresentada na questão 1, o nível de significância do teste é $0,07031 + 0,01758 + 0,001953 = 0,090$, que é maior do que $0,05$, de forma que o teste é liberal. A estatística de teste é $B = \sum_{i=1}^{n} \psi_i$, em que $\psi_i = 1$, se $Z_i > 0$, e $\psi_i = 0$, se $Z_i < 0$, sendo que $Z_i = X_i - \theta_0$. Rejeitamos H_0 se $B_{obs} \geq b_{1-\alpha} = 7$; caso contrário, não rejeitamos. Aplicando a transformação $z_i = x_i - \theta_0$ obtemos
-0,09, 0,02, 0,41, -0,05, 0,56, 0,32, 0,28 0,18 e 0,25,
de modo que $B_{obs} = 7$ e a decisão é rejeitar H_0. A um nível de significância de 9%, conclui-se que a dose

efetiva mediana ultrapassa 0,5.
(b) Como $B_{\text{obs}} = 7$, valor-$p = \sum_{j=7}^{9} \text{P}(B = j)$, sendo que $B \sim \text{binomial}(9, 1/2)$. Aproveitando um cálculo no item (a), obtemos valor-p = 0,090. A um nível de significância de 5%, conclui-se que a dose efetiva mediana não ultrapassa 0,5.

Obs. O teste com região crítica $B_{obs} \geq 8$ tem nível de significância $0,01758+0,001953 = 0,020 < \alpha = 0,05$.

3. Em um experimento com culturas de milho foram realizadas contagens de quatro tipos de plantas (M1, M2, M3 e M4), obtendo-se 670, 230, 238 e 62. De acordo com a teoria, a distribuição dos tipos de plantas deve ser nas proporções 9:3:3:1. (a) A teoria é compatível com os dados coletados? (b) O que você pode afirmar sobre o valor-p do teste realizado?
Solução. De acordo com a teoria, as probabilidades dos quatro tipos de plantas, M1 a M4, são $\theta_{10} = 9/16$, $\theta_{20} = \theta_{30} = 3/16$ e $\theta_{40} = 1/16$.
(a) Os dados foram coletados de uma amostra com $n = 1200$. Logo, as frequências esperadas sob H_0 ($e_i = n \times \theta_{i0}$, $i = 1, \ldots, 4$) são $e_1 = 675$, $e_2 = e_3 = 225$ e $e_4 = 75$. Para os dados observados, a estatística de teste de Pearson tem valor
$$Q_{obs} = \frac{(670-675)^2}{675} + \frac{(230-225)^2}{225} + \frac{(238-225)^2}{225} + \frac{(62-75)^2}{75} = 3,15,$$
correspondendo a $4 - 1 = 3$ g.l. (H_0 é simples). Com um nível de significância de 5%, consultando a tabela da nota 2 notamos que para $a = 0,05$ o valor crítico é maior do que 4,64, de modo que $Q_{obs} = 3,15$ é menor do que o valor crítico. Não rejeitamos a hipótese de que a distribuição dos tipos de plantas segue as proporções 9:3:3:1.
(b) Como $2,95 < Q_{obs} = 3,15 < 3,66$, pela tabela da nota 2 vemos que $0,30 <$ valor-$p < 0,40$.

4. Estudantes realizaram uma prova e obtiveram escores 95, 80, 40, 52, 60, 80, 82, 58, 65 e 50. Afirma-se que a função distribuição acumulada do escore (/100) é $F_0(x) = 0$, se $x < 0$, $F_0(x) = x^2(3 - 2x)$, se $x \in [0, 1]$, e $F_0(x) = 1$, se $x > 1$. (a) Com base nestes dados, o que pode ser concluído sobre a afirmação? (b) O que você pode afirmar sobre o valor-p do teste realizado?
Solução. Deve ser testada $\text{H}_0 : F(x) = F_0(x)$ para todo x contra $\text{H}_1 : F(x) \neq F_0(x)$ para pelo menos um x, em que $F_0(x)$ está apresentada no enunciado. Adotamos um nível de significância de 5%. Como a hipótese nula é simples, o teste será realizado com a estatística de Kolmogorov-Smirnov. As $n = 10$ observações (escore/100) em ordem crescente são

0,40, 0,50, 0,52, 0,58, 0,60, 0,65, 0,80, 0,80, 0,82 e 0,95.

A função distribuição empírica $S_n(x)$ tem incremento igual a $1/n = 1/10$ em cada um dos valores observados, exceto que para o escore 0,80 o incremento é igual a 2/10. Os elementos necessários ao cálculo da estatística de teste D_n estão apresentados na Tabela 1, com $\varepsilon = 0,05$. A estatística D_n é calculada como

Tabela 1: Cálculo da estatística D_n na questão 4.

x	$S_n(x)$	$F_0(x)$	$\mid S_n(x) - F_0(x) \mid$	$\mid S_n(x - \varepsilon) - F_0(x) \mid$
0,40	0,1	0,352	0,252	0,352
0,50	0,2	0,500	0,300	$\boxed{0,400}$
0,52	0,3	0,530	0,230	0,330
0,58	0,4	0,619	0,219	0,319
0,60	0,5	0,648	0,148	0,248
0,65	0,6	0,718	0,118	0,218
0,80	0,8	0,896	0,096	0,296
0,82	0,9	0,914	0,014	0,114
0,95	1,0	0,993	0,007	0,093

o valor máximo dos elementos das duas últimas colunas da Tabela 1, resultando em $D_{n,obs} = 0,400$. Consultando a tabela da nota 1 obtemos $D_{n,1-\alpha} = 0,409$. Como $D_{n,\text{obs}} < D_{n,1-\alpha}$, não rejeitamos a hipótese

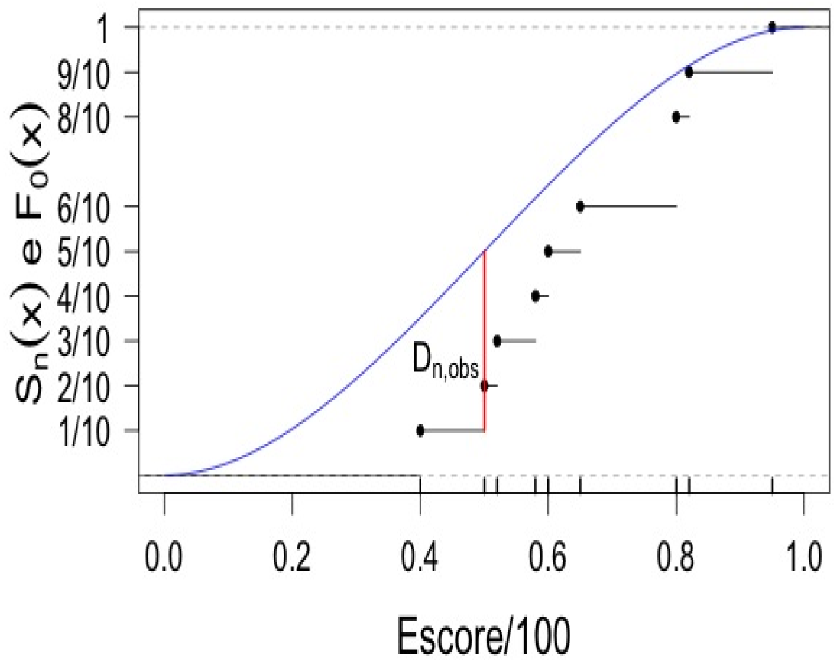

Figura 1: Funções distribuição e $D_{n,\text{obs}}$ na questão 4.

NOTA 1.

d	0,323	0,369	$D_{n,obs} = 0,400$	0,409	0,489
$P(D_{10} \geq d)$	0,20	0,10	valor-p	0,05	0,01

nula. A um nível de significância de 5%, não podemos negar a afirmação de que os escores seguem a distribuição especificada. A Figura 1 mostra uma solução gráfica do problema.
(b) A partir da tabela da nota 1, percebemos que $0,369 < D_{n,\text{obs}} < 0,409$. Portanto, podemos afirmar que $0,05 <$ valor-$p < 0,10$.

NOTA 1. Abaixo são apresentados alguns valores de probabilidades da cauda direita da distribuição da estatística de Kolmogorov-Smirnov para $n = 10$.
NOTA 2. Abaixo são apresentados alguns valores de w tais que $P(W \geq w) = a$ e para alguns graus de liberdade (g), sendo que $W \sim \chi^2_g$.

NOTA 2.

	a				
g	0,50	0,40	valor-p	0,30	0,20
2	1,39	1,83		2,41	3,22
3	2,37	2,95	$Q_{obs} = 3,15$	3,66	4,64
4	3,36	4,04		4,88	5,99