USP – ICMC – SME0810 - Métodos Não Paramétricos

Prova 1 – 2023

S O L U Ç Ã O

1. A sequência de símbolos

   A, A, A, B, B, A, A, A, A, B, A, A, B, A, A, A, A, B, B, B, B

   indica dias com temperatura máxima acima (A) ou abaixo (B) da média histórica de um certo período. Pode ser afirmado que as temperaturas diárias se afastam da média histórica de forma independente?

   Solução. Os números de símbolos A e B são $n_1 = 13$ e $n_2 = 8$, respectivamente, de modo que $n = n_1 + n_2 = 21$. Na sequência de símbolos notamos quatro corridas de A e quatro corridas de B (vide abaixo), de modo que o número observado de corridas é $R_{\text{obs}} = 8$.

   $$\underbrace{A, A, A}_{1}, \overbrace{B, B}^{1}, \underbrace{A, A, A, A}_{2}, \overbrace{B}^{2}, \underbrace{A, A}_{3}, \overbrace{B}^{3}, \underbrace{A, A, A, A}_{4}, \overbrace{B, B, B, B}^{4}$$

   Supondo independência, usamos a Nota 2. Temos $\lambda^* = n_1/n = 13/21 = 0,619$ e $Z_{\text{obs}} = \{R_{\text{obs}} - 2n\lambda^*(1 - \lambda^*)\}/\{2\sqrt{n}\lambda^*(1 - \lambda^*)\} = -0,881$. Utilizando a aproximação pela distribuição normal padrão, com 5% de significância temos $|z_{\text{crit}}| = 1,96$. Como $|Z_{\text{obs}}| < |z_{\text{crit}}|$, não rejeitamos a hipótese de independência. Com 5% de significância, concluímos que as temperaturas diárias se afastam da média histórica de forma independente.

2. Um total de $n = 50$ lotes de $m = 13$ itens cada foi coletado em uma linha de produção. As quantidades de defeitos e lotes são apresentados abaixo. Pode ser afirmado que o número de defeitos segue a distribuição binomial?

Defeitos (x)	0	1	2	3	4	5	6 ou mais
Lotes (f)	10	24	10	4	1	1	0

   Solução. O número de defeitos é denotado por X. Seguindo o enunciado, será testada a hipótese composta $X \sim \text{binomial}(13, \pi)$, com um parâmetro ($\pi$) a ser estimado. Utilizando a Nota 3 e a tabela acima, obtemos a estimativa da probabilidade π dada por $\widehat{\pi} = \sum_{j=0}^{m} x_j \times f_j/(n \times m) = \sum_{j=0}^{13} x_j \times f_j/650 = 65/650 = 0,100$. Aplicamos o teste qui-quadrado de Pearson de bondade de ajuste. Na tabela abaixo, as probabilidades (`Prob.`) foram calculadas com a função massa de probabilidade da distribuição binomial, ou seja, $\binom{13}{j} \times 0,100^j \times 0,900^{13-j}$, para $j = 0, 1, \ldots, 13$. A coluna das frequências esperadas estimadas (`e^`) é calculada como $n\times$ `Prob.` = 50 $\times$ `Prob.`.

```
---------------------------------------------------------
Defeitos Amostras (o) Prob.     e^ (e^ - o)^2 / e^
---------------------------------------------------------
       0           10 0,254 12,709             0,578
       1           24 0,367 18,358             1,734
```

```
        2           10 0,245 12,239           0,409
        3            4 0,100  4,986           0,195
        4            1 0,028  1,385           0,107
        5            1 0,006  0,277           1,887
  6 ou mais          0 0,001  0,046           0,046
-------------------------------------------------
          Total     50 1,000 50,000           4,956
-------------------------------------------------
```

O teste é realizado com os resultados da tabela acima em que o número de categorias é $k = 7$ (ou seja, sem agrupar o número de defeitos para que a frequência esperada estimada `e^` seja pelo menos 5). Na tabela acima vemos que $Q_{\text{obs}} = 4,956$. O número de graus de liberdade (g.l.) é g.l. $= k - 1 - 1 = 5$. Consultando a tabela da Nota 4, vemos que $4,35 < 4,956 < 6,63$, de modo que o valor-p correspondente a $Q_{\text{obs}} = 4,956$ é tal que $0,25 <$ valor-$p < 0,50$, ou seja, valor-$p > 0,05$. Logo, fixando o nível de significância do teste em 5%, concluímos que a distribuição binomial$(13, p)$ faz um bom ajuste aos dados.

3. Nove laboratórios participaram de um estudo com o objetivo de verificar se a dose efetiva mediana de uma certa droga ultrapassa 0,5 mg/l. Os dados coletados foram

   0,41, 0,52, 0,91, 0,45, 1,06, 0,82, 0,78, 0,68 e 0,75.

   (a) Formule as hipóteses a testar e indique dois testes que poderiam ser aplicados.
   (b) Apresente uma solução com base no valor-p de um teste apropriado.

   Solução. (a) Deve ser tetada $\text{H}_0 : \theta = \theta_0$ contra $\text{H}_1 : \theta > \theta_0$, 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). Em princípio, os testes de Wilcoxon e do sinal poderiam ser usados.
   (b) A solução está baseada no teste do sinal. A estatística de teste é $B = \sum_{i=1}^{n} I(Z_i > 0)$, em que $n = 9$ e $Z_i = X_i - \theta_0$. 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 $B_{\text{obs}} = 7$. Com a distribuição $B \sim$ binomial$(9, 1/2)$, o valor-p é dado por

   $$\text{valor-}p = \sum_{j=7}^{9} \text{P}(B = j) = \frac{1}{2^9} \sum_{j=7}^{9} \binom{9}{j} = 0,0898,$$

   ou seja, valor-$p > 0,05$. Logo, a um nível de significância de 5%, conclui-se que a dose efetiva mediana não ultrapassa 0,5.

4. Pode ser afirmado que as observações

   -0,53 -0,40 0,52 -0,27 -0,24 -0,07 -0,76 -0,58 -0,49 0,34

   foram amostradas de uma distribuição uniforme$(-1, 1)$?

   Solução. Pelo enunciado, a distribuição é contínua e a hipótese nula

   $$\text{H}_0 : X \sim \text{uniforme}(-1, 1)$$

   é simples. Utilizamos o teste de Kolmogorov-Smirnov. A função densidade é $f(x) = 1/2$, se $x \in (-1, 1)$, e $f(x) = 0$, caso contrário. Para $x \in (-1, 1)$, temos

   $$F_0(x) = \int_{-1}^{x} \frac{1}{2} du = \frac{1}{2} u\Big|_{-1}^{x} = \frac{x + 1}{2}.$$

Para $x \leq -1$, temos $F_0(x) = 0$ e para $x \geq 1$, temos $F_0(x) = 1$. Os valores das $n = 10$ observações são diferentes, de modo que a função distribuição empírica F_n apresenta saltos de altura igual a $1/n = 1/10$ em cada diferente valor observado. Os cálculos necessários são organizados na tabela abaixo.

```
--------------------------------------------------
    x Fn(x) F0(x) |F0(x) - Fn(x)| |F0(x) - Fn(x-)|
--------------------------------------------------
-0,76   0,1 0,120           0,020            0,120
-0,58   0,2 0,210           0,010            0,110
-0,53   0,3 0,235           0,065            0,035
-0,49   0,4 0,255           0,145            0,045
-0,40   0,5 0,300           0,200            0,100
-0,27   0,6 0,365           0,235            0,135
-0,24   0,7 0,380           0,320            0,220
-0,07   0,8 0,465           0,335            0,235
 0,34   0,9 0,670           0,230            0,130
 0,52   1,0 0,760           0,240            0,140
--------------------------------------------------
```

Calculando o valor máximo das duas últimas colunas da tabela acima, obtemos $D_{\text{obs}} = 0,335$. Consultando a tabela da Nota 1, vemos que $0,323 < 0,335 < 0,369$, de modo que o valor-p correspondente a $D_{\text{obs}} = 0,335$ é tal que 0,10 < valor-p < 0,20, ou seja, valor-$p > 0,05$. Logo, fixando o nível de significância do teste em 5%, não rejeitamos a afirmação de que as observações foram amostradas de uma distribuição uniforme$(-1, 1)$.

Nota 1. Abaixo são apresentados alguns valores de probabilidades da cauda direita da distribuição da estatística D_n de Kolmogorov-Smirnov para $n = 10$.

d	0,323	0,369	0,409	0,489
$\mathrm{P}(D_{10} \geq d)$	0,20	0,10	0,05	0,01

Nota 2. Supondo independência da sequência, a distribuição aproximada de $Z = \{R - 2n\lambda^*(1-\lambda^*)\}/\{2\sqrt{n}\lambda^*(1-\lambda^*)\}$ é $N(0, 1)$, sendo que R é número total de corridas e $\lambda^* = n_1/n$.

Nota 3. Na questão 2, um estimador para a probabilidade de sucesso da distribuição binomial é dado por $\sum_{j=0}^{m} x_j \times f_j/(n \times m)$.

Nota 4. No corpo da tabela abaixo são apresentados alguns valores de w tais que $\mathrm{P}(W \geq w) = p$ para diferentes graus de liberdade (g.l.), sendo que $W \sim \chi^2_{\text{g.l.}}$.

	p			
g.l.	0,50	0,25	0,10	0,05
4	3,36	5,39	7,78	9,49
5	4,35	6,63	9,24	11,07
6	5,35	7,84	10,64	12,59