Unidade 2 – Testes para uma amostra

Teste de Aleatoriedade ou de Iterações

Tem como objetivo verificar se uma amostra extraída de uma população é, realmente, aleatória.

Para aplicar o teste devemos ter os dados dicotomizados, para que possamos realizar uma contagem do número “r” de iterações.

Pressuposições:

Os dados analisados consistem de uma sequência de observações, registradas na ordem de suas ocorrências, os quais podem ser classificados dentro de dois tipos mutuamente exclusivos, onde:

1. n é o tamanho da amostra;
2. n_1 é o número de observações do tipo “A”;
3. n_2 é o número de observações do tipo “B”.

Hipóteses:

H_o: a ordem dos símbolos é aleatória;

H_1: a ordem dos símbolos não é aleatória.

Procedimento

Dispor as observações na ordem de sua ocorrência e contar o número de iterações de “a” e de “b”, ou seja, o valor de “r”; Isso se baseia na ordem ou seqüência em que os escores individuais foram obtidos originalmente, sendo que uma iteração é definida como uma sucessão de símbolos idênticos que aparecem seguidos e precedidos por símbolos diferentes (ou por nenhum símbolo).

Por exemplo, suponha que nossos dados sejam codificados em

AAAAABBAAABAABBBBBBBBAA

Começando pelo primeiro elemento “A”, enquanto temos somente “A” o sucedendo, estamos na primeira iteração. No momento em que mudamos para “B” passaremos a segunda iteração que acontecerá até que apareça um próximo “A” e mudaremos para a terceira iteração e assim sucessivamente. No caso:

A A A A A	B B	A A A	B	A A	B B B B B B B B	A A
1	2	3	4	5	6	7

Cada cor acima representa uma iteração. Observe os pontos de mudança de iterações. Quando saímos da sequencia de símbolos iguais e começa um símbolo diferente, alteramos a iteração. Nesse caso temos $r = 7$ iterações e a partir dessa informação podemos testar se essa sequência de “A” e “B” é aleatória ou não.

Quando os dados já se apresentam dicotomizados por natureza, fica fácil obter o valor de r. No caso em que temos dados numéricos, devemos aplicar uma regra para atribuir dois símbolos: atribuímos “A” aos valores maiores que a mediana e “B” aos valores menores que a mediana. A partir daí conseguimos contar as iterações normalmente.

Regra de Decisão

Depende do tamanho dos grupos n_1 e n_2; se n_1 e n_2 são menores ou iguais a 20, recorrer à tabela do teste (disponível no moodle).

Essa tabela tem duas partes, FI e FII, que são os limites para o valor de “r”. Se $FI < r < FII$, ou seja, estiver dentro dos limites, não rejeitamos a hipótese nula. Do contrário se $r \leq FI$ ou $r \geq FII$, rejeitamos Ho. Observe que a tabela é simétrica, logo não temos critério em escolher quem é n_1 ou n_2.

Exemplo: Verifique a aleatoriedade da sequencia abaixo:

AAAAABBAAABAABBBBBBBBAA

Resolução:

$$\begin{cases} H_0: \text{ é aleatório} \\ H_1: \text{não é aleatório} \end{cases}$$

Devemos encontrar n_1 (numero de A's), n_2 (número de B´s) e o valor de r (número de iterações).

A A A A A B B A A A B A A B B B B B B B B A A

1	2	3	4	5	6	7

Assim, $n_1 = 12$ e $n_2 = 11$ e r = 7 iterações.

Na tabela, cruzando os valores n_1 e n_2 encontramos os limites.

F_I

n_1 \ n_2	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
2											2	2	2	2	2	2	2	2	2
3					2	2	2	2	2	2	2	2	2	3	3	3	3	3	3
4				2	2	2	3	3	3	3	3	3	3	3	4	4	4	4	4
5			2	2	3	3	3	3	3	4	4	4	4	4	4	4	5	5	5
6		2	2	3	3	3	3	4	4	4	4	5	5	5	5	5	5	6	6
7		2	2	3	3	3	4	4	5	5	5	5	5	6	6	6	6	6	6
8		2	3	3	3	4	4	5	5	5	6	6	6	6	6	7	7	7	7
9		2	3	3	4	4	5	5	5	6	6	6	7	7	7	7	8	8	8
10		2	3	3	4	5	5	5	6	6	7	7	7	7	8	8	8	8	9
11		2	3	4	4	5	5	6	6	7	7	7	8	8	8	9	9	9	9
12	2	2	3	4	4	5	6	6	7	7	7	8	8	8	9	9	9	10	10
13	2	2	3	4	5	5	6	6	7	7	8	8	9	9	9	10	10	10	10
14	2	2	3	4	5	5	6	7	7	8	8	9	9	9	10	10	10	11	11
15	2	3	3	4	5	6	6	7	7	8	8	9	9	10	10	11	11	11	12
16	2	3	4	4	5	6	6	7	8	8	9	9	10	10	11	11	11	12	12
17	2	3	4	4	5	6	7	7	8	9	9	10	10	11	11	11	12	12	13
18	2	3	4	5	5	6	7	8	8	9	9	10	10	11	11	12	12	13	13
19	2	3	4	5	6	6	7	8	8	9	10	10	11	11	12	12	13	13	13
20	2	3	4	5	6	6	7	8	9	9	10	10	11	12	12	13	13	13	14

F_{II}

n_1 \ n_2	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
2																	
3																	
4		9	9														
5	9	10	10	11	11												
6	9	10	11	12	12	13	13	13	13								
7		11	12	13	13	14	14	14	14	15	15	15					
8		11	12	13	14	14	15	15	16	16	16	16	17	17	17	17	17
9			13	14	14	15	16	16	16	17	17	18	18	18	18	18	18
10			13	14	15	16	16	17	17	18	18	18	19	19	19	20	20
11			13	14	15	16	17	17	18	19	19	19	20	20	20	21	21
12			13	14	16	16	17	18	19	19	20	20	21	21	21	22	22
13				15	16	17	18	19	19	20	20	21	21	22	22	23	23
14				15	16	17	18	19	20	20	21	22	22	23	23	23	24
15				15	16	18	18	19	20	21	22	22	23	23	24	24	25
16					17	18	19	20	21	21	22	23	23	24	25	25	25
17					17	18	19	20	21	22	23	23	24	25	25	26	26
18					17	18	19	20	21	22	23	24	25	25	26	26	27
19					17	18	20	21	22	23	23	24	25	26	26	27	27
20					17	18	20	21	22	23	24	25	25	26	27	27	28

Verificamos então que FI = 7 e F2 = 18.

Como r = 7 e é **menor ou igual** a FI que também é 7**, rejeitamos H_0.** Não existem evidências para afirmarmos que exista aleatoriedade.

Se n_1 e n_2 são maiores que 20, utilizamos a padronização abaixo:

$$Z = \frac{r - \mu_r}{\sigma_r} = \frac{r - \left(\frac{2n_1 n_2}{n_1 + n_2} + 1\right)}{\sqrt{\frac{2n_1 n_2 (2n_1 n_2 - n_1 - n_2)}{(n_1 + n_2)^2 (n_1 + n_2 - 1)}}}$$

Para decidir pela aceitação ou rejeição de H_o, considerar os valores tabelados da variável $Z \approx N(0,1)$.

Regra de decisão:

Se o valor de p, associado ao valor de r, não for superior a α, rejeita-se H_o. Caso contrário, aceita-se H_o.

Exemplo: Verifique a aleatoriedade da seguinte amostra constituída pelos resultados de 40 lançamentos de uma moeda onde K representa cara e C coroa.

{K, K, C, C, K, C, K, K, C, K, C, C, K, K, K, C, K, K, C, K, K, C, C,

K, C, K, K, C, K, C, C, K, K, C, K, K, K, C, C, K}