Tabelas Bidimensionais

Unidade I

Parte 5

Inferência para Pequenas Amostras

Quando:

- Tamanho da amostra é pequeno
- Existem células com frequências pequenas ou nulas

⇓

Usar DISTRIBUIÇÕES EXATAS para fazer inferência.

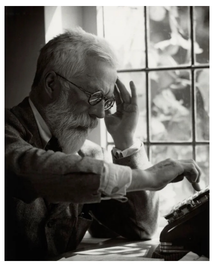
- ***Exemplo 1:***

  Em um estudo sobre tratamentos para curar infecções graves, um tratamento em teste e um controle são comparados para determinar se as taxas de resposta favorável são as mesmas. Doze pacientes selecionados aleatoriamente para receber o tratamento em teste e outros 6 pacientes para receber placebo.

Tratamento	Resposta ao Tratamento		Total
	Favorável	Desfavorável	
Teste	10	2	12
Controle	2	4	6
Total	12	6	18

Exemplo 2:

Uma colega de Fisher na Estação Experimental de Rothamsted próxima de Londres sustenta que, quando toma chá, ela pode distinguir se leite ou chá foi colocado primeiro na xícara. Para testar sua afirmação, Fisher delineou um experimento no qual ela testava oito xícaras de chá. Em quatro xícaras foi colocado primeiro leite, e nas ourtras quatro foi colocado primeiro chá. Foi dito a ela que haviam quatro xícaras de cada tipo, de modo que ela pode tentar selecionar as quatro em que foram colocadas leite primeiro. As xícaras eram apresentadas a ela em ordem aleatória.



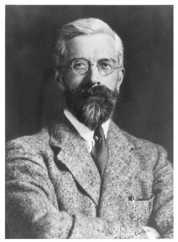
Ronald A . Fisher
(1890-1962)

Adicionado Primeiro	Acho que serviu primeiro		Total
	Leite	Chá	
Leite	3	1	4
Chá	1	3	4
Total	4	4	8

- Desejamos testar as seguintes hipóteses:

$H_0)\ \theta = 1$ $(A\ adivinhação\ da\ colega\ de\ Fisher\ independe\ da\ real\ ordem\ de\ colocação\ do\ chá.)$

$H_0)\ \theta > 1$ $(Reflete\ a\ afirmação\ da\ colega\ de\ Fisher, predizendo\ uma\ associação\ positiva\ entre\ a\ verdadeira\ ordem\ de\ colocação\ do\ chá\ e\ a\ adivinhação.)$



Ronald A . Fisher
(1890-1962)

⇓

TESTE EXATO DE FISHER

Teste Exato de Fisher

Em tabelas 2x2 temos:

X \ Y	sucesso	insucesso	Total
1	n_{11}	n_{12}	n_{1+}
2	n_{21}	n_{22}	n_{2+}
Total	n_{+1}	n_{+2}	n

Fixados os totais, dado n_{11}, as demais frequências estão determinadas.

X \ Y	sucesso	insucesso	Total
1	n_{11}	$n_{1+} - n_{11}$	n_{1+}
2	$n_{+1} - n_{11}$	$n_{2+} - n_{21}$	n_{2+}
Total	n_{+1}	n_{+2}	n

Lembrando que em uma distribuição HIPERGEOMÉTRICA temos:

N* objetos** *r* ***sucessos

N - r ***insucessos***

Seleciona-se n e sendo

X: número de sucessos nas n observações.

$$P(X = x) = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}$$

Podemos pensar como:

n **observações** n_{+1} ***sucessos***

$n_{+2} = n - n_{+1}$ ***insucessos***

Selecionada uma amostra de tamanho n_{1+} do nível 1 de X, temos que :

n_{11}: número de sucessos em n_{1+} obser-vações do nível 1 de X

tem distribuição amostral HIPERGEOMÉTRICA

Teste Exato de Fisher

- Hipóteses:

$$H_0)\ \theta = 1$$

$$H_1)\ \theta > 1$$

- Estatística do Teste:

n_{11}: *número de sucessos em n_{1+} observações do nível 1 de **X***

tem distribuição amostral HIPERGEOMÉTRICA

$$P(n_{11}) = \frac{\binom{n_{+1}}{n_{11}}\binom{n-n_{+1}}{n_{1+}-n_{11}}}{\binom{n}{n_{1+}}} = \frac{n_{+1}!\, n_2!\, n_{1+}!\, n_{2+}!}{n!\, n_{11}!\, n_{21}!\, n_{12}!\, n_{22}!}$$

X \ Y	1	2	Total	n_{11}	P(n11)
1	n_{11}	n_{12}	4		
2	n_{21}	n_{22}	4		
Total	4	4	8		
	1	2			
1	0	4	4		
2	4	0	4	0	0,01428571
	4	4	8		
	1	2			
1	1	3	4		
2	3	1	4	1	0,22857143
	4	4	8		
	1	2			
1	2	2	4		
2	2	2	4	2	0,51428571
	4	4	8		
	1	2			
1	3	1	4		
2	1	3	4	3	0,22857143
	4	4	8		
	1	2			
1	4	0	4		
2	0	4	4	4	0,01428571
	4	4	8		

- Estatística do teste: $n_{11}=3$
- Distribuição Amostral Exata:

Distribuição Exata da Estatística n_{11}

n_{11}	Probabilidade
0	0,014285714
1	0,228571429
2	0,514285714
3	0,228571429
4	0,014285714
Total	**1,000000000**

$$p - valor = P(3) + P(4) = 0{,}229 + 0{,}014 = 0{,}243$$

onde

n_{11} é

No caso da hipótese alternativa ser:

$$H_1)\ \theta \neq 1$$

O ***p-valor exato*** é definido como a soma das probabilidades de todos os resultados $\boldsymbol{y}$ para a frequência da primeira célula (*estatística do teste*) para os quais $\boldsymbol{P(y) \leq P(n_{11})}$, sendo que $\boldsymbol{n_{11}}$ é a frequência observada.

Exemplo:

$$p - valor = P(0) + P(1) + P(3) + P(4) = 0{,}014 + 0{,}229 + 0{,}229 + 0{,}014 = 0{,}486.$$

Mid p-value

- Por causa da estatística do teste ser discreta, não é usualmente possível alcançar o nível de significância exatamente.
- O teste é dito ser conservativo, visto que a real taxa de erro é menor que a pretendida.
- Para diminuir a tendência do teste a ser conservador, recomendamos usar o ***p-valor médio*** (***mid p-value***).

mid p-value é igual a metade da probabilidade do resultado observado, mais a probabilidade de resultados mais extremos.

Exemplo:

$$mid\ p - valor = \frac{p(3)}{2} + P(4) = \frac{0{,}229}{2} + 0{,}014 = 0{,}129$$

Distribuição Exata da Estatística χ^2

- Estatística: χ^2

n_{11}	Probabilidade	χ^2
0	0,014	8
1	0,229	2
2	0,514	0
3	0,229	2
4	0,014	8

A estatística pode assumir apenas três valores distintos, portanto sua distribuição altamente discreta está longe da distribuição contínua qui-quadrado.

Distribuição Exata da Estatística χ^2

χ^2	Probabilidade
0	0,514285714
2	0,457142857
8	0,028571429
Total	**1,000000000**

A tabela observada tem $X^2 = 2$ e o p-valor é igual a

$$\boldsymbol{p - valor} = P(\chi^2 \geq 2) = 0{,}458 + 0{,}028 = 0{,}486.$$