ME111 - Laboratório de Estatística

Aula 7 - Teste Exato de Fisher

Profa. Larissa Avila Matos

Teste Exato de Fisher

- O teste exato de Fisher é baseado no cálculo da distribuição de probabilidade das frequências da tabela. Contudo, isso não é possível na situação das tabelas com margens livres ou com uma margem fixa e outra livre porque a probabilidade de uma dada distribuição das frequências é função de parâmetros de valor desconhecido.

- Fisher (1934) propôs que a distribuição de probabilidade das frequências de qualquer um destes tipos de tabelas sejam substituídas pela probabilidade da distribuição das mesmas frequências, considerando tabelas com duas margens fixas, ou seja, uma distribuição de probabilidade hipergeométrica para a única frequência de valor livre (independente).

		Resposta		Total
		S (Sucesso)	F (Fracasso)	
Tratamento	1	n_{11}	n_{12}	$n_{11} + n_{12}$
	2	n_{21}	n_{22}	$n_{21} + n_{22}$
Total		$n_{11} + n_{21}$	$n_{12} + n_{22}$	n

- $n = n_{11} + n_{12} + n_{21} + n_{22}$.
- $\hat{p_1} = \frac{n_{11}}{n_{11}+n_{12}}$ é a proporção amostral de sucessos no tratamento 1.
- $\hat{p_2} = \frac{n_{21}}{n_{21}+n_{22}}$ é a proporção amostral de sucessos no tratamento 2.
- H_0 : a probabilidade de sucesso é a mesma em cada um dos tratamentos, isto é, $p_1 = p_2$; em que p_i é a verdadeira (populacional) proporção de sucesso do tratamento i.

- Estatística do teste: $x = \hat{p_1} - \hat{p_2}$.
- Precisamos então calcular a distribuição de probabilidade da estatística do teste, sob H_0.

- Para a tabela acima (como visto na aula passada), temos que

$$\mathbb{P}(Y = y) = \frac{\begin{pmatrix} n_{11} + n_{12} \\ y \end{pmatrix} \begin{pmatrix} n_{21} + n_{22} \\ n_{11} + n_{21} - y \end{pmatrix}}{\begin{pmatrix} n \\ n_{11} + n_{21} \end{pmatrix}}, \quad y = 0, \ldots, n_{11},$$

  onde Y é a frequência da célula (1,1).

- Ou seja $Y \sim$ Hipergeométrica.

Distribuição Hipergeométrica

- População dividida em duas características;
- Extrações casuais sem reposição;
- N objetos;
- r têm a característica A;
- $N - r$ têm a característica B;
- Um grupo de n elementos é escolhido ao acaso, dentre os N possíveis, sem reposição.

Queremos calcular a probabilidade de que este grupo de n elementos contenha x elementos com a característica A.

Elemento escolhido	Característica A	Característica B	Total
Sim	x	$n - x$	n
Não			$N - n$
Total	r	$N - r$	N

- Seja X a v.a. que representa o número de elementos com a característica A dentre os n escolhidos ao acaso.

Então dizemos que X segue uma distribuição Hipergeométrica com parâmetros N, n, r, ou seja, $X \sim Hip(N, n, r)$.

A probabilidade de se observar x é dada por:

$$\mathbb{P}(X = x) = \frac{\begin{pmatrix} r \\ x \end{pmatrix} \begin{pmatrix} N - r \\ n - x \end{pmatrix}}{\begin{pmatrix} N \\ n \end{pmatrix}}, \quad 0 \leq x \leq \min\{r, n\}.$$

Uma senhora toma chá

- Uma senhora inglesa desafiou seus colegas cientistas dizendo que era capaz de distinguir a ordem de adicionamento (quem foi primeiro) de leite ou chá na mistura leite+chá. Sir Ronald Fisher decidiu então realizar o seguinte experimento:
  - 8 xícaras: 4 com chá colocado antes do leite e 4 com leite antes do chá.
  - As oito xícaras foram apresentadas em ordem aleatória para a senhora, mas a informação de que eram 4 de cada tipo foi passada a ela.
  - A senhora deveria provar a bebida das oito xícaras e escolher as 4 xícaras que acreditava estar com chá primeiro.
  - Verificou-se quantas dentre as 4 xícaras ela escolheu corretamente.
- Quais os resultados possíveis do experimento?

		Palpite		Total
		Chá 1^o	Leite 1^o	
Tratamento	Chá 1^o	n_{11}		4
	Leite 1^o			4
Total		4	4	8

- Marginais fixas, conhecimento de uma casela determina as demais.

Teste de hipótese

- H_0 : A senhora não consegue distinguir entre chá ou leite primeiro e escolhe ao acaso durante o experimento, ou seja, não existe associação do palpite com a ordem dos ingredientes.

- H_1 : A senhora consegue distinguir, ou seja, existe associação do palpite com a ordem dos ingredientes.

- Estatística do teste: Total de acertos (Y).

- Para decidir, precisamos primeiramente da distribuição de probabilidade da estatística do teste quando H_0 é verdadeira.

- Queremos a probabilidade sob H_0 de encontrarmos um Y com valor igual ou superior ao n_{11} da tabela observada no experimento?
- Como calculamos esta probabilidade?
- E se não soubéssemos calcula-la?
- Tendo esta probabilidade em mãos, como decidimos a favor de H_0?

- Experimento - Tabela observada:

		Palpite		Total
		Chá 1^o	Leite 1^o	
Tratamento	Chá 1^o	3	1	4
	Leite 1^o	1	3	4
Total		4	4	8

- Como decidir se rejeitamos ou não H_0 de acordo com a estatística do teste observada?

- Como decidir se rejeitamos a hipótese de que a senhora não consegue distinguir os chás, sendo que ela acertou, por exemplo, 3? Se ela tivesse acertado todas as 4 xícaras? Seria por pura sorte? Ou ela tem algum conhecimento?

- Calculamos a probabilidade de um valor igual ou mais extremo ao da estatística do teste observada (valor-de-p). Mais extremo: mais evidência contra H_0.

- Se o valor-de-p obtido é bem pequeno, por exemplo, $0,01$, isto quer dizer que se H_0 é verdadeira, então seria incomum obter uma amostra com os resultados como o observado.

- Um valor-de-p muito baixo traz fortes evidências contra H_0.

- Portanto,

$$p - valor = \mathbb{P}(Y = 3) + \mathbb{P}(Y = 4) = 0.2285714 + 0.0142857 = 0.2428571$$

Teste Exato de Fisher - Simulação

```r
x = NULL
B = 1000
n1=n2=4
for (i in 1:B){
s = sample(c(1,1,1,1,2,2,2,2))
x[i] = sum(s[1:n1]==1)}

table(x)/B
```

```
x
    0     1     2     3     4
0.010 0.214 0.518 0.247 0.011
```

```r
x.obs = 3
p.valor = mean(ifelse(x >= x.obs,1,0))
p.valor
```

```
[1] 0.258
```

Teste Exato de Fisher no R

```r
TeaTasting = matrix(c(3, 1, 1, 3), nrow = 2,
                    dimnames = list(Guess = c("Milk", "Tea"),
                                    Truth = c("Milk", "Tea")))
TeaTasting
```

```
      Truth
Guess  Milk Tea
  Milk    3   1
  Tea     1   3
```

```r
# "greater" teste unilateral
fisher.test(t(TeaTasting), alternative = "greater")
```

```

    Fisher's Exact Test for Count Data

data:  t(TeaTasting)
p-value = 0.2429
alternative hypothesis: true odds ratio is greater than 1
95 percent confidence interval:
 0.3135693       Inf
sample estimates:
odds ratio
  6.408309
```

Exercício

- De uma maneira geral os doentes psiquiátricos podem ser classificados em psicóticos e neuróticos. Um psiquiatra realiza um estudo sobre os sintomas suicidas em duas amostras de 20 doentes de cada grupo. Os resultados tabelados são

		Tipo de Doente		
		Psicóticos	Neuróticos	Total
Sintomas Suicidas	Presente	2	6	8
	Ausente	18	14	32
Total		20	20	40

- H_0 :A nossa hipótese é de que a proporção de psicóticos com sintomas suicidas é igual a proporção de neuróticos com estes sintomas.

- Qual a probabilidade sob H_0 de encontrarmos um Y com valor igual ou inferior ao n_{11} da tabela observada no experimento?