

MAT02010 - Tópicos Avançados em Estatística II

Inferência Causal em Experimentos Aleatorizados

Rodrigo Citton P. dos Reis
citton.padilha@ufrgs.br

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
DEPARTAMENTO DE ESTATÍSTICA

Porto Alegre, 2019

Ausência efeito é plausível?

Os protocolos agressivo e menos agressivo são equivalentes?

- ▶ Estudo **ProCESS**: suponha que você está inicialmente inclinado a acreditar que o protocolo agressivo não impediu e não causou mortes entre os $I = 885$ pacientes quando contrastado com o protocolo menos agressivo.
- ▶ Os resultados do estudo forçam você a revisar essa crença?
 - ▶ $m = 439$ pacientes tratados \rightsquigarrow taxa de mortalidade intra-hospitalar de 21,0%.
 - ▶ $I - m = 446$ pacientes controles \rightsquigarrow taxa de mortalidade intra-hospitalar de 18,2%.
- ▶ As duas taxas de mortalidade acima se referem a **pacientes diferentes**, e talvez pacientes diferentes tenham taxas de mortalidade diferentes simplesmente porque são pacientes diferentes, com infecções diferentes e diferentes estados de saúde.

Os protocolos agressivo e menos agressivo são equivalentes?

- ▶ Este é um estudo aleatorizado.
 - ▶ Os dois grupos são amostras aleatórias da população de $n = 885$ pacientes.
 - ▶ Os dois grupos são similares (aleatorização garante balanceamento das covariáveis).
- ▶ A diferença entre uma taxa de mortalidade de 21,0% e uma taxa de mortalidade de 18,2% pode ser devida **ao acaso**, ao lançamento de uma moeda que atribuiu um paciente ao protocolo agressivo e o próximo ao protocolo menos agressivo?

Os protocolos agressivo e menos agressivo são equivalentes?

```
# Ausência de efeito de tratamento
r_T <- r_C <- c(rep(1, 30), rep(0, 70))

# Atribuição de tratamento
set.seed(2306)
Z <- rbinom(n = 100, size = 1, prob = 0.5)

# Desfecho observado
R <- ifelse(Z == 1, r_T, r_C)
round(prop.table(table(Z, R), margin = 1) * 100, 2)
```

```
##      R
## Z      0      1
## 0 67.39 32.61
## 1 72.22 27.78
```

Os protocolos agressivo e menos agressivo são equivalentes?

- ▶ Estamos perguntando se taxas de mortalidade de 21,0% e 18,2% poderiam facilmente surgir por acaso **se a hipótese nula de Fisher de ausência (nenhum) efeito** de tratamento fosse verdadeira.
- ▶ A hipótese H_0 de Fisher de ausência efeito afirma que $r_{T_i} = r_{C_i}$ para todo i , $i = 1, \dots, I$, em que $I = 885$ no estudo ProCESS.
 - ▶ Isso diz que o paciente i pode morrer ou não, mas a sobrevivência do paciente i sob tratamento agressivo, r_{T_i} , é a mesma que a sobrevivência do paciente i sob o tratamento menos agressivo, r_{C_i} .
- ▶ Equivalentemente, a hipótese H_0 de Fisher de ausência efeito afirma que a diferença de efeito, $\delta_i = r_{T_i} - r_{C_i}$, é zero para cada paciente, $H_0 : \delta_i = 0, i = 1, \dots, I$.

Os protocolos agressivo e menos agressivo são equivalentes?

- ▶ É perfeitamente concebível que a hipótese de Fisher de ausência efeito seja verdadeira, o que implica $0 = \bar{\delta} = \bar{r}_T - \bar{r}_C$, mas $\hat{r}_T - \hat{r}_C = 2,8\%$ porque \hat{r}_T foi estimado em pacientes no grupo de tratamento agressivo, enquanto \hat{r}_C foi estimado em diferentes pacientes no grupo de tratamento menos agressivo.
- ▶ Se H_0 fosse verdade, a moeda lançada que designou um paciente para tratamento e outro para controle produziu facilmente taxas de mortalidade de 21,0% e 18,2% nos dois grupos de tratamento?

Notação tabular

Table 3.1. In-hospital mortality in the ProCESS Trial with general notation

Treatment group	<i>In-hospital mortality</i>		Total	Death rate (%)
	<i>Death</i> $R_i = 1$	<i>Other</i> $R_i = 0$		
Aggressive, $Z_i = 1$	$T = 92$	347	$m = 439$	$\hat{r}_T = 21.0$
Less aggressive, $Z_i = 0$	81	365	$I - m = 446$	$\hat{r}_C = 18.2$
Total	173	712	$I = 885$	

- ▶ Fisher usou T como a **estatística de teste**.
 - ▶ $T = \sum_{i=1}^I Z_i R_i = \sum_{i=1}^I Z_i r_{Ti} = m \times \hat{r}_T$.
- ▶ Existem duas diferenças entre a primeira e a segunda linha da Tabela:
 1. (grupo de) pessoas diferentes; similares por aleatorização, mas ainda assim, diferentes;
 2. as pessoas da primeira linha receberam o protocolo agressivo e as pessoas da segunda linha receberam o tratamento menos agressivo.

Notação tabular

- ▶ Repetindo, $\hat{r}_T = 21,0\%$ e $\hat{r}_C = 18,2\%$ podem diferir por dois motivos: **(i)** são calculados a partir de pessoas diferentes e **(ii)** é possível que o tratamento agressivo tenha efeitos diferentes do tratamento menos agressivo, $r_{Ti} \neq r_{Ci}$ para algum paciente i .
- ▶ A hipótese nula de Fisher de ausência de efeito, H_0 , é uma negação da segunda explicação: diz que $r_{Ti} = r_{Ci}$ para cada paciente i , então $\hat{r}_T = 21,0\%$ e $\hat{r}_C = 18,2\%$ diferem apenas devido à maneira como as moedas escolhem as pessoas para os dois grupos.
- ▶ **Se a hipótese nula** de Fisher de ausência efeito, H_0 , **fosse verdadeira**, qual a chance de que o lançamento de uma moeda iria, por puro acaso, colocar $T = 92$ ou mais mortes no grupo tratamento agressivo?
 - ▶ Essa chance, essa probabilidade, é chamada de **valor- P** ou **nível de significância**.

O ensaio de uniformidade

O que é um ensaio de uniformidade

- ▶ Em um ensaio de uniformidade, a distinção entre parcelas tratadas e controle foi mantida, mas todas as parcelas foram tratadas da mesma maneira, com o mesmo fertilizante e o mesmo inseticida.
- ▶ Usando ensaios de uniformidade, os pesquisadores aprenderam empiricamente quanto os grupos tratado e controle poderiam diferir quando não havia efeito do tratamento, porque todas as parcelas foram tratadas da mesma maneira.
 - ▶ Como todos os participantes de um ensaio de uniformidade recebem o mesmo tratamento, por exemplo, o controle, as duas linhas da tabela análoga à Tabela 3.1 registram o tratamento Z_i atribuído e a resposta ao controle r_{C_i} .
 - ▶ No ensaio de uniformidade, **as duas linhas** da tabela análoga à Tabela 3.1 **diferem** apenas porque se referem a **pessoas diferentes**, porque os dois grupos de pessoas receberam o mesmo tratamento.

O ensaio de uniformidade e a hipótese de Fisher

- ▶ Se o estudo ProCESS fosse um ensaio de uniformidade, em que todos os indivíduos recebesse o protocolo menos agressivo, então $R_i = r_{C_i}$ e $T^* = \sum_{i=1}^I Z_i R_i = \sum_{i=1}^I Z_i r_{C_i}$.
 - ▶ **Não vemos** T^* porque é o número de mortes que teria ocorrido entre as pessoas do grupo de tratamento agressivo se elas fossem tratadas usando o protocolo menos agressivo..
- ▶ Se a hipótese de Fisher de ausência de efeito, H_0 , fosse verdadeira, então $r_{T_i} = r_{C_i}$ para todo i , $i = 1, \dots, I = 885$, e assim $T = T^*$.
 - ▶ Portanto, testar a hipótese de nenhum efeito consiste em perguntar se T exibe o tipo de comportamento que esperaríamos de T^* em um teste de uniformidade, um teste sem efeito de tratamento, ou se precisamos acreditar em $r_{T_i} \neq r_{C_i}$ para algum paciente i se quisermos entender o comportamento de T .

O ensaio de uniformidade e a hipótese de Fisher

- ▶ A quantidade $A = \sum_{i=1}^I Z_i \delta_i = T - T^*$ é chamada de **efeito atribuível**: é o aumento líquido de mortes no grupo tratado causado pelo recebimento de tratamento e não de controle.
 - ▶ **Não vemos** A , pois não vemos T^* .
 - ▶ Se a hipótese de nenhum efeito de Fisher fosse verdadeira, $r_{T_i} = r_{C_i}$ para cada paciente i , $i = 1, \dots, I = 885$ e $A = 0$.
- ▶ Se a hipótese de Fisher de ausência de tratamento é verdadeira, então vemos o efeito causal (r_{T_i}, r_{C_i}) para todo i , pois $r_{T_i} = r_{C_i}$.
 - ▶ Neste caso, a Tabela 3.1 seria o resultado de um ensaio de uniformidade, pois $r_{T_i} = r_{C_i}$.
- ▶ Se a Tabela 3.1 tivesse vindo de um estudo de uniformidade (**ausência de efeito**), qual é a chance de haver 92 ou mais mortes no grupo de tratamento agressivo?

Testando nenhum efeito: um exemplo pequeno

Possíveis atribuições de tratamento e suas probabilidades

- ▶ Imagine um ensaio clínico aleatorizado com $I = 8$ pacientes dos quais $m = 4$ são escolhidos aleatoriamente para o protocolo agressivo no estudo ProCESS, e os demais $I - m = 4$ pacientes recebem o protocolo menos agressivo (controle).
- ▶ O paciente i tem $Z_i = 1$ se atribuído ao protocolo agressivo e $Z_i = 0$ se atribuído ao protocolo menos agressivo e $4 = m = Z_1 + \dots + Z_8$.

Possíveis atribuições de tratamento e suas probabilidades

- ▶ Existem 70 maneiras de escolher $m = 4$ pacientes para tratamento de $I = 8$ pacientes ($70 = \binom{8}{4}$).
- ▶ Sempre que você escolhe quatro dos oito pacientes para tratamento, define quatro dos Z_i igual a 1 e os quatro restantes como 0.
 - ▶ Se você selecionou os pacientes 1, 2, 3 e 5 para tratamento, define $Z_1 = Z_2 = Z_3 = 1$, $Z_4 = 0$, $Z_5 = 1$, $Z_6 = Z_7 = Z_8 = 0$.
- ▶ Portanto, existem 70 maneiras de definir quatro dos Z_i igual a 1 e os quatro restantes Z_i iguais a 0.

Possíveis atribuições de tratamento e suas probabilidades

Table 3.2. Abbreviated table of 70 possible treatment assignments

Assignment	Probability	Treated	Treatment indicators							
			Z_1	Z_2	Z_3	Z_4	Z_5	Z_6	Z_7	Z_8
1	1/70	1, 2, 3, 4	1	1	1	1	0	0	0	0
2	1/70	1, 2, 3, 5	1	1	1	0	1	0	0	0
				⋮						
44	1/70	2, 3, 6, 8	0	1	1	0	0	1	0	1
				⋮						
70	1/70	5, 6, 7, 8	0	0	0	0	1	1	1	1

Os resultados do ensaio

- ▶ Com probabilidade $1/70$ escolhemos (por sorteio) a configuração 44 da Tabela 3.2 ($Z_1 = 0, Z_2 = 1, Z_3 = 1, Z_4 = 0, Z_5 = 0, Z_6 = 1, Z_7 = 0, Z_8 = 1$).
 - ▶ Não há nada de especial com esta atribuição de tratamento; escolhemos por sorte!
- ▶ Os pacientes foram então tratados da maneira ditada pelos números aleatórios e a mortalidade intra-hospitalar, R_i , foi registrada.
- ▶ Para cada paciente i , registramos $R_i = 1$ se o paciente morreu ou $R_i = 0$ se o paciente sobreviveu.
- ▶ Imagina-se que o ensaio tenha resultados extremos para a sobrevivência; especificamente, $R_1 = 0, R_2 = 1, R_3 = 1, R_4 = 0, R_5 = 0, R_6 = 1, R_7 = 0, R_8 = 1$, produzindo a Tabela a seguir.

Os resultados do ensaio

Table 3.3. In-hospital mortality in the small hypothetical experiment

<i>Treatment group</i>	<i>In-hospital mortality</i>		<i>Total</i>	<i>Death rate (%)</i>
	<i>Death</i> $R_i = 1$	<i>Other</i> $R_i = 0$		
Treated, $Z_i = 1$	$T = 4$	0	$m = 4$	$\hat{r}_T = 100$
Control, $Z_i = 0$	0	4	$I - m = 4$	$\hat{r}_C = 0$
Total	4	4	$I = 8$	

- ▶ A Tabela 3.3 pode ser devido ao acaso?
 - ▶ É uma diferença dramática, mas há apenas $I = 8$ pacientes.
 - ▶ O padrão da Tabela 3.3 poderia surgir facilmente apenas por acaso, se o tratamento não tivesse efeito?

A lógica dos testes de hipóteses

- ▶ Para testar a hipótese nula de ausência de efeito tratamento, supomos, provisoriamente, apenas por uma questão de argumento, que isso é verdade.
- ▶ Supor que a hipótese nula seja verdadeira ao testar essa hipótese não tem nada a ver com acreditar que a hipótese nula é verdadeira.
 - ▶ Hipóteses nulas não são o tipo de coisa em que você acredita ou não, e, em qualquer caso, a **crença não tem papel aqui**.
 - ▶ Testar uma hipótese nula pergunta se os dados que vimos - aqui, a Tabela 3.3 - **fornecem fortes evidências** de que a hipótese nula é falsa ou, alternativamente, se os dados que vimos fornecem pouca orientação sobre se a hipótese nula é verdadeira ou falsa.

A lógica dos testes de hipóteses

- ▶ A lógica dos testes de hipóteses pergunta: se a hipótese nula fosse verdadeira, veríamos uma tabela como a Tabela 3.3?
 - ▶ Obviamente, poderíamos ver uma tabela como a Tabela 3.3 (é uma possibilidade lógica), mas é altamente improvável?
- ▶ Se a hipótese nula de nenhum efeito do tratamento fosse verdadeira em um experimento aleatorizado com $m = 4$ pessoas escolhidas aleatoriamente de $I = 8$ pessoas, qual é a chance de vermos $T = 4$ ou mais mortes no grupo tratado na Tabela 3.3?

A distribuição da estatística de teste T quando a hipótese nula é verdadeira

- ▶ Se a hipótese nula de nenhum efeito for verdadeira, então $r_{T_i} = r_{C_i}$, então para todos os $i = 1, \dots, 8$ pacientes, $R_i = r_{C_i}$.
- ▶ Se a hipótese fosse verdadeira, saberíamos pelos **dados observados** que $r_{C_1} = 0$, $r_{C_2} = 1$, $r_{C_3} = 1$, $r_{C_4} = 0$, $r_{C_5} = 0$, $r_{C_6} = 1$, $r_{C_7} = 0$, $r_{C_8} = 1$.
- ▶ Além disso, se a hipótese fosse verdadeira, saberíamos que alterar os tratamentos que os pacientes receberam não alteraria suas respostas.
- ▶ Em geral, se a hipótese nula fosse verdadeira, então $T = \sum_{i=1}^8 Z_i r_{C_i}$, portanto, sabemos o que T teria sido em cada uma das 70 situações da Tabela 3.2.

A distribuição da estatística de teste T quando a hipótese nula é verdadeira

Table 3.4. Abbreviated null distribution of T in the small hypothetical randomized experiment

Assignment	Probability	T	Responses under the hypothesis of no effect							
			$R_1 = 0 = r_{C1}$	$R_2 = 1 = r_{C2}$	$R_3 = 1 = r_{C3}$	$R_4 = 0 = r_{C4}$	$R_5 = 0 = r_{C5}$	$R_6 = 1 = r_{C6}$	$R_7 = 0 = r_{C7}$	$R_8 = 1 = r_{C8}$
			Treatment indicators							
			Z_1	Z_2	Z_3	Z_4	Z_5	Z_6	Z_7	Z_8
1	1/70	2	1	1	1	1	0	0	0	0
2	1/70	2	1	1	1	0	1	0	0	0
				⋮						
44	1/70	4	0	1	1	0	0	1	0	1
				⋮						
70	1/70	2	0	0	0	0	1	1	1	1

A distribuição da estatística de teste T quando a hipótese nula é verdadeira

- ▶ A Tabela 3.5 é a distribuição da estatística de teste T quando a hipótese nula de nenhum efeito é verdadeira e desempenha um papel importante no teste da hipótese nula.
 - ▶ É a distribuição de T sob a hipótese nula em um experimento aleatorizado no qual $m = 4$ de $I = 8$ pacientes foram escolhidos aleatoriamente para tratamento, com respostas observadas $R_1 = 0$, $R_2 = 1$, $R_3 = 1$, $R_4 = 0$, $R_5 = 0$, $R_6 = 1$, $R_7 = 0$, $R_8 = 1$.
- ▶ A Tabela 3.5 é construída a partir de duas premissas e uma atividade.
 1. A primeira premissa é que atribuímos tratamentos usando números verdadeiramente aleatórios; essa premissa produz probabilidades, $1/70$.
 2. A segunda premissa é que a hipótese de nenhum efeito de tratamento é verdadeira, para que possamos deduzir o valor de T em todas as atribuições de tratamento possíveis a partir das **respostas que observamos**.
 3. A atividade é contar em uma tabela com 70 linhas, em nosso caso, a Tabela 3.4.

A distribuição da estatística de teste T quando a hipótese nula é verdadeira

Table 3.5. The distribution of T when the null hypothesis is true

t	Possible values t of T					Total
	0	1	2	3	4	
Number of rows with $T = t$	1	16	36	16	1	70
$\Pr(T = t)$	1/70	16/70	36/70	16/70	1/70	1
$\Pr(T \geq t)$	70/70	69/70	53/70	17/70	1/70	

- ▶ $T \geq 4$ é muito demais para ser “devido ao acaso”.
- ▶ Como observamos $T \geq 4$ e $T \geq 4$ é improvável quando a hipótese nula de nenhum efeito for verdadeira, dizemos que a Tabela 3.3 constitui evidência bastante forte contra a hipótese nula em nosso pequeno experimento aleatório.

Comparando tratamentos no estudo ProCESS

Quão grande é o efeito?

Avisos

- ▶ **Para casa:** Ler o Capítulo 4 do livro do Paul R. Rosenbaum.
 - ▶ Ler o Capítulo 5 do livro de Guido Imbens e Donald Rubin.
 - ▶ Ler o artigo “Statistics and Causal Inference” do Paul Holland.
- ▶ **Próxima aula:** Discussão do Capítulo 4 do livro do Paul R. Rosenbaum.
 - ▶ A discussão deste capítulo será condizida pelos estudantes da turma.
 - ▶ Ver Cap. 1 do livro “**Statistics: a guide to the unknown**”.

Por hoje é só!

