Lecture 21

Ciprian M Crainiceanu

Department of Biostatistics
Johns Hopkins Bloomberg School of Public Health
Johns Hopkins University

December 5, 2013

Table of contents

1. Introduce Fisher's exact test
2. Illustrate Monte Carlo version of test

Fisher's exact test

not assuming normality

- Fisher's exact test is "exact" because it guarantees the α rate, regardless of the sample size
- Example, chemical toxicant and 10 mice

	Tumor	None	Total
Treated	4	1	5
Control	2	3	5
Total	6	4	

- p_1 = prob of a tumor for the treated mice
- p_2 = prob of a tumor for the untreated mice

Continued

score

个人理解：当SS很小的时候、才会用 fisher

- $H_0 : p_1 = p_2 = p$
- Can't use Z or χ^2 because SS is small
- Don't have a specific value for p

Fisher's exact test

- Under the null hypothesis every permutation is equally likely
- observed data

Treatment	:	T	T	T	T	T	C	C	C	C	C
Tumor	:	T	T	T	T	N	T	T	N	N	N

- permuted break the connection [illegible]. fix margin

Treatment	:	T	C	C	T	C	T	T	C	T	C
Tumor	:	N	T	T	N	N	T	T	T	N	T

5T 5C

6T 4N

- Fisher's exact test uses this null distribution to test the hypothesis that $p_1 = p_2$

Hyper-geometric distribution

- X number of tumors for the treated
- Y number of tumors for the controls
- $H_0 : p_1 = p_2 = p$
- Under H_0
  - $X \sim \text{Binom}(n_1, p)$
  - $Y \sim \text{Binom}(n_2, p)$
  - $X + Y \sim \text{Binom}(n_1 + n_2, p)$

Continued

x		n_1
$z-x$		n_2
z		n_1+n_2

$\Rightarrow$ 只看第一列

$$\frac{C_{n_1}^{x} \cdot C_{n_2}^{z-x}}{C_{n_1+n_2}^{z}}$$

$$P(X = x \mid X + Y = z) = \frac{\binom{n_1}{x}\binom{n_2}{z-x}}{\binom{n_1+n_2}{z}}$$

This is the hypergeometric pmf

Proof

$$P(X = x) = \binom{n_1}{x} p^x (1-p)^{n_1 - x}$$

$$P(Y = z - x) = \binom{n_2}{z - x} p^{z-x} (1-p)^{n_2 - z + x}$$

$$P(X + Y = z) = \binom{n_1 + n_2}{z} p^z (1-p)^{n_1 + n_2 - z}$$

Continued

$$
\begin{aligned}
P(X = x \mid X + Y = z) &= \frac{P(X = x, X + Y = z)}{P(X + Y = z)} \\
&= \frac{P(X = x, Y = z - x)}{P(X + Y = z)} \\
&= \frac{P(X = x)P(Y = z - x)}{P(X + Y = z)}
\end{aligned}
$$

Plug in and finish off yourselves

Fisher's exact test

- More tumors under the treated than the controls
- Calculate an *exact* P-value
- Use the conditional distribution = hypergeometric
- Fixes both the row and the column totals
- Yields the same test regardless of whether the rows or columns are fixed
- Hypergeometric distribution is the same as the permutation distribution given before

Tables supporting H_a

- Consider $H_a : p_1 > p_2$ "one-sided"
- P-value requires tables as extreme or more extreme (under H_a) than the one observed
- Recall we are fixing the row and column totals
- Observed table

Table 1 =

4	1	5
2	3	5
6	4	

两个 margin 都固

- More extreme tables in favor of the alternative

Table 2 =

5	0	5
1	4	5
6	4	

margin fixed

Calculations

$$P(\text{Table 1}) = P(X = 4 | X + Y = 6) = \frac{\binom{5}{4}\binom{5}{2}}{\binom{10}{6}} = 0.238$$

$$P(\text{Table 2}) = P(X = 5 | X + Y = 6) = \frac{\binom{5}{5}\binom{5}{1}}{\binom{10}{6}} = 0.024$$

P-value $= 0.238 + 0.024 = 0.262$

R code

$$\begin{array}{cc} & \text{Tumor} \\ \text{Treat} & \begin{pmatrix} 4 & 1 \\ 2 & 3 \end{pmatrix} \end{array}$$

$\hat{p}_1 = \frac{4}{5}$

$\hat{p}_2 = \frac{2}{5}$

$H_0: p_1 = p_2$

```
dat <- matrix(c(4, 1, 2, 3), 2)
fisher.test(dat, alternative = "greater")

-------------------output-----------------
        Fisher's Exact Test for Count Data

data:  dat
p-value = 0.2619
alt hypoth: true odds ratio  is greater than 1
95 percent confidence interval:
 0.3152217       Inf
sample estimates:
odds ratio
  4.918388
```

Notes

- Two sided p-value = 2×one sided P-value
  (There are other methods which we will not discuss)
- P-values are usually large for small n
- Doesn't distinguish between rows or columns
- The common value of p under the null hypothesis is called a nuisance parameter
- Conditioning on the total number of successes, $X + Y$, eliminates the nuisance parameter, p
- Fisher's exact test guarantees the type I error rate
- Exact unconditional P-value

$$\sup_p P(X/n_1 > Y/n_2; p)$$

Monte Carlo

- Observed table $X = 4$

```
Treatment  : T T T T T C C C C C
Tumor      : T T T T N T T N N N
```

- Permute the second row

打散

```
Treatment  : T T T T T C C C C C
Tumor      : T N T N T T N N T T
```

- Simulated table $X = 3$
- Do over and over
- Calculate the proportion of tables for which the simulated $X \geq 4$
- This proportion is a Monte Carlo estimate for Fisher's exact P-value