Oct. 18

Contingency Tables: Part Two

Cross-Product ratio

Prospective Design

	$Y=1$	$Y=2$
$X=1$	π_{11}	$1-\pi_{11}$
$Y=2$	π_{21}	$1-\pi_{21}$

X & Y "unrelated" means

$$P(Y=1 \mid X=1) = P(Y=1 \mid X=2)$$

$$\pi_{11} - \pi_{11}\pi_{21} = \pi_{21} - \pi_{11}\pi_{21}$$

$$\Rightarrow \pi_{11}(1-\pi_{21}) = \pi_{21}(1-\pi_{11})$$

$$\Leftrightarrow \frac{\pi_{11}(1-\pi_{21})}{\pi_{21}(1-\pi_{11})} = 1$$

CROSS-PRODUCT RATIO

$$\parallel$$

$$\theta$$

$$\frac{\text{Product of Main diagonal}}{\text{Product of off-diagonal}}$$

Retrospective Design

	$Y=1$	$Y=2$
$X=1$	π_{11}	π_{12}
$X=2$	$1-\pi_{11}$	$1-\pi_{12}$

Unrelated means $P(X=1 \mid Y=1) = P(X=1 \mid Y=2)$

$$\pi_{11} - \pi_{11}\pi_{12} = \pi_{12} - \pi_{11}\pi_{12}$$

$$\pi_{11}(1-\pi_{12}) = \pi_{12}(1-\pi_{11})$$

$$\Longleftrightarrow \frac{\pi_{11}(1-\pi_{12})}{\pi_{12}(1-\pi_{11})} = 1$$

$$\Longleftrightarrow \theta = 1$$

Cross-Sectional Design

	$Y=1$	$Y=2$	
$X=1$	π_{11}	$a-\pi_{11}$	a
$X=2$	$b-\pi_{11}$	$1-a-(b-\pi_{11})$ $=1-a-b+\pi_{11}$	$1-a$
	b	$1-b$	1

Re-parameterize

$$(\pi_{11}, \pi_{12}, \pi_{21}) \longleftrightarrow (\pi_{11}, a, b)$$

Show X & Y independent iff

$$\pi_{11} = ab$$

$\pi_{11} = ab \Leftrightarrow$

	$Y=1$	$Y=2$	
$X=1$	ab	$a-ab = a(1-b)$	a
$X=2$	$b-ab = b(1-a)$	$1-a-(b-ab)$ $1-a-b+ab$ $=(1-a)(1-b)$	$1-a$
	b	$1-b$	1

Independent

Showing X & Y independent

$$\iff \theta = 1$$

Independence

ab	$a(1-b)$	a
$(1-a)b$	$(1-a)(1-b)$	$1-a$
b	$1-b$	

$$\Rightarrow \theta = \frac{ab(1-a)(1-b)}{(1-a)b\,a(1-b)} = 1$$

Going in other direction & NOT assuming independence, table is

π_{11}	$a-\pi_{11}$	a
$b-\pi_{11}$	$1-a-b+\pi_{11}$	$1-a$
b	$1-b$	1

Let $\theta = 1$

$$1 = \frac{\pi_{11} - \pi_{11}a - \pi_{11}b + \pi_{11}^2}{ab - a\pi_{11} - b\pi_{11} + \pi_{11}^2}$$

$$\Longleftrightarrow\ ab\ \cancel{- a\pi_{11} - b\pi_{11} + \pi_{11}^2} = \pi_{11}\ \cancel{- a\pi_{11} - b\pi_{11} + \pi_{11}^2}$$

$$ab = \pi_{11} \Longleftrightarrow X \,\&\, Y \text{ independent}$$

Cross-product ratio <u>is</u> the Odds Ratio

<u>Odds</u> If $P(A) = \pi$

"Odds" of $A = \dfrac{\pi}{1-\pi}$ ("to one")

- If $P(A) = \pi = \frac{2}{3}$ odds $= \dfrac{2/3}{1/3} = 2$ to 1
- If $\pi = 9/10$, odds $\dfrac{9/10}{1/10} = 9$ to 1
- If $\pi = \frac{2}{7}$

  odds $= \dfrac{2/7}{5/7} = \dfrac{2}{5} = 0.4$ to one

Instead of saying 2 to 5.

Instead of odds 1 to 4, say 0.25 to one

Odds can be <u>any</u> non-neg #

Conditional Odds: Use Conditional Probabilities

Odds Ratio

$$\frac{\text{Odds of } Y=1 \text{ Given } X=1}{\text{Odds of } Y=1 \text{ Given } X=2}
\qquad
\begin{array}{c|c|c|}
 & Y=1 & Y=2 \\
\hline
X=1 & \pi_{11} & \pi_{12} \\
\hline
X=2 & \pi_{21} & \pi_{22} \\
\hline
\end{array}$$

$$\frac{\pi_{11} / \cancel{\pi_{11}+\pi_{12}} \;\Big/\; \pi_{12} / \cancel{\pi_{11}+\pi_{12}}}{\dfrac{\pi_{21}}{\cancel{\pi_{21}+\pi_{22}}} \;\Big/\; \dfrac{\pi_{22}}{\cancel{\pi_{21}+\pi_{22}}}} = \frac{\pi_{11}\,\pi_{22}}{\pi_{12}\,\pi_{21}} = \theta$$

Exercise to show for other 2 study designs

Cross-Product ratio is the odds Ratio.

Meaningful for larger tables

Dept		Admit	Reject
	A		
	B		
Dept	C		
	D		
	F		

Odds Ratio odds of admit to Dept A vs odds of admit for Dept C

If hear "Chances of death before 50 are 4 times as great for smokers"

OR it could be "relative risk."

Relative Risk $= \dfrac{P(Y=1 \mid X=1)}{P(Y=1 \mid X=2)}$ 9

$= 1 = \Theta = 1$ if X & Y are unrelated, but otherwise they can tell a somewhat different story.

Fisher's exact test

Permutation test (Fisher's Design of Experiments, 1935)

Experimental study, prospective, with random assignment of units to conditions.

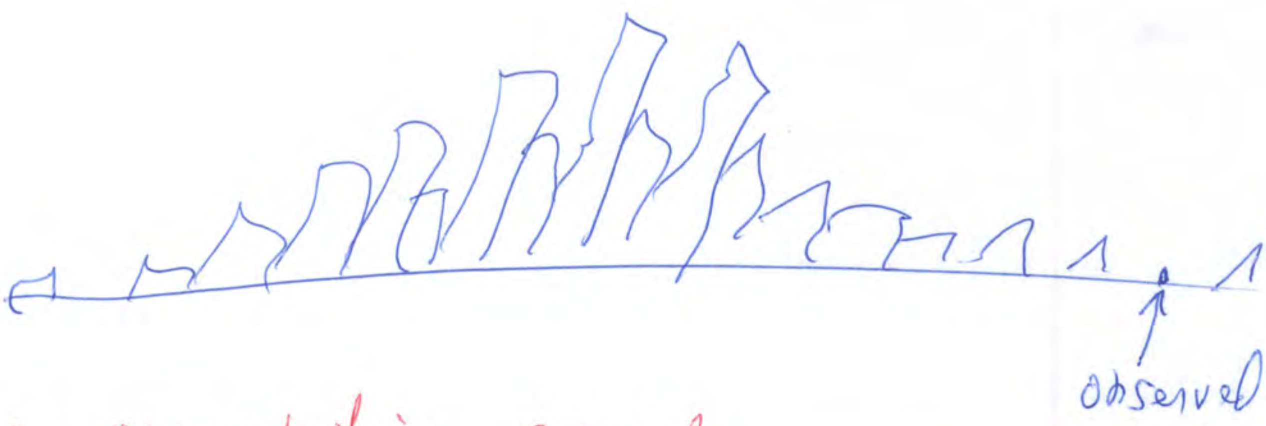
Under H_0: Treatment has no effect at all

* Process of producing y values is unspecified, except nothing to do with experimental condition
* y values are what they are
* Only reason for diff between conditions is RANDOM ASSIGNMENT.

Notice no ~~pretense~~ pretence of Random Sampling.

Permutation Distribution

- Pick a test statistic
- Under H_0, all ways of distributing the y values into conditions are =ally likely.
- Compute test statistic for each re-arrangement of data
- Relative frequencies are the permutation distribution of test statistic



the permutation p-value is the proportion of values in the permutation distribution that equal or exceed the value from the original un-scrambled data, in the direction of the alternative hypothesis.

Simple

Distribution-free

Large sample not required

No pretense of random sampling

Applies to obs studies as a test of independence.

But

Lots of ways to re-arrange the data

Control: 8 plots of land

6 Exp conditions, 4 plots each

$n = 32$

There are $\binom{32}{8\ 4\ 4\ 4\ 4\ 4\ 4} = \dfrac{32!}{8!\,4!^6}$

ways to put y values into boxes

12.5

Calculate the F statistic for each re-arrangement of the data

$$\frac{32!}{8!\,4!\,4!\,4!\,4!\,4!\,4!} = 34,149,454,710,484,113,000,000$$

- That's a big number.
- Maybe we can distribute the computation among lots of computers.
- World population is approximately 7.51 billion.
- That's $34149454710484113/7510 \approx 4.547 \times 10^{12}$ calculations per person.
- If they all had computers and could do one test every 0.01 seconds,
- It would take around 1,441.9 years to finish the job.

14

13

Fisher said

Statistical methods for research workers, 1936

> *Actually, the statistician does not carry out this very tedious process but his conclusions have no justification beyond the fact they could have been arrived at by this very elementary method.*

See Cox and Reid (2000) *The Theory of the Design of Experiments* for the research literature.

Fisher's exact test

	$Y=1$	$Y=2$	
$X=1$	$x = n_{11}$	$a-x$	a
$X=2$	$b-x$	$n-a-b+x$	$n-a$
	b	$n-b$	n

$$\hat{\theta} = \frac{n_{11}/n \; n_{22}/n}{n_{21}/n \; n_{12}/n} = \frac{n_{11}\, n_{22}}{n_{21}\, n_{12}}$$

Sample cross-product ratio

Goes up & down with x

Test statistic is sample odds ratio $\longleftrightarrow x$

	$Y=1$	$Y=2$	
$X=1$	x	$a-x$	a
$X=2$	$b-x$	$n-a-b+x$	$n-a$
	b	$n-b$	n

Under H_0, all re-arrangements of X & Y are equally likely

What is $P_r(n_{11}=x)$?

$$\frac{\binom{n}{x,\ \ a-x,\ \ b-x,\ \ n-a-b+x}}{\binom{n}{a}\cdot\binom{n}{b}}$$

16

$$= \frac{\dfrac{\cancel{n!}}{x!\,(a-x)!\,(b-x)!\,(n-a-b+x)!}}{\dfrac{\cancel{n!}}{a!\,(n-a)!}\;\dbinom{n}{b}}$$

$$= \frac{\dfrac{a!}{x!\,(a-x)!} \cdot \dfrac{(n-a)!}{(b-x)!\,(n-a-(b-x))!}}{\dbinom{n}{b}}$$

$$= \frac{\dbinom{a}{x}\dbinom{n-a}{b-x}}{\dbinom{n}{b}}$$ ✓

Hypergeometric: Class of 60 G & 40 B, choose 10 w/o replacement

$x = \#$ Girls

$$P(x) \quad \frac{\dbinom{60}{x}\dbinom{40}{10-x}}{\dbinom{100}{10}}$$

oct 20

	$Y=1$	$Y=2$	
$X=1$	x	$a-x$	a
$X=2$	$b-x$	$n-a-b+x$	$n-a$
	b	$n-b$	n

Under H_0

$$P(n_{11}=x) = \frac{\binom{a}{x}\binom{n-a}{b-x}}{\binom{n}{b}} = \frac{\binom{b}{x}\binom{n-b}{a-x}}{\binom{n}{a}}$$

over what possible values of x ?

$x \geq 0$, $x \leq a$, $x \leq b$,

$n-a-b+x \geq 0 \iff x \geq a+b-n$, so

$$\mathrm{Max}(0, a+b-n) \leq x \leq \mathrm{Min}(a, b)$$

One tailed p-value is probability of getting greater (less) than or = to observed n_{11}

2-tailed p-value is sum of all probabilities (in both tails) less than or = to prob of observed n_{11}.