Assignment 4 due next Thurs!

Review of last day
- 2 x 2 Contingency Tables

- 2 variables to relate
  - categorical variables - binary
  - factors -

data

		Column Factor		
		colcat$_1$	colcat$_2$	
Row factor	row cat$_1$	.	.	r_1
	row cat$_2$	.	.	r_2
		c_1	c_2	n

taken Column Factor - response
Row Factor - explanatory

consider p_1 = Prob(in column cat 1 | row cat 1)

p_2 = Prob(in column cat 1 | row cat 2)

Pain trtmt example

Row factor - Treatment variable
- IRS . active
- control . placebo

Col. factor - Pain relief
- Yes or no

interested in p_1 = Prob(Relief | IRS)
p_2 = Prob(Relief | control)

Inference for $p_1 - p_2$

Confidence intervals
→ $(\hat{p}_1 - \hat{p}_2) \pm z_{\alpha/2}\ s.e.(\hat{p}_1 - \hat{p}_2)$

large sample
smaller samples. can
use a continuity correction
→ makes s.e. bigger
→ default in R: prop.test

Test procedure. $H_0: p_1 - p_2 = 0$

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{s.e.(\hat{p}_1 - \hat{p}_2)}$$

$\frac{\text{est} - \text{hyp. val}}{s.e.}$

→ different fla from C.I
– two flavours – with & without continuity correction

can also use Pearson's χ^2
– with & without C.C.

$\chi^2 = (Z)^2$ – in prop.test → uses χ^2

Simplified, desert island approach

$$s.e.(\hat{p}_1 - \hat{p}_2) = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

CI and Z test

- quick & dirty calc.

- inference . statistical inference

→ statement about pattern in population

- scientific inference

→ interpretation in practically relevant perspective

Two studies -

Randomized trial

- IRS shows evidence of clinical potential

→ Juvenile delinquents eyeglasses

Accept the statistical inf

that Juvenile Delinquents
aren't wearing glasses

- ? what does it mean
  - why is this the case ??

is there a causal relationship
?? what lies beneath the
the pattern of association?

Causal hypotheses
- lack of glasses
  leads to delinquency

but maybe poverty → delinquency
→ lack of glasses

→ poverty is potential
confounder in this
study

→ non-delinquents - richer families
- could assess incomes
of all subjects to
examine this possibility

→ observational study
→ no control over factors.

c.f. randomized study
↙ - control over
~~treatment~~ factor
treatment

observational - naturalistic
no intervention

→ prone to confounding
→ randomized studies protect against confounding

Study design features
- experiments - assign "treatment"
  - experimental conditions
  → randomized experiments
- observational studies

Temporality ← flow of time
- what came first
→ RCT → assign treatments / apply treatments → observe outcome

- prospective study
  - follow subjects over time

Study 2. Juvenile delinq's

- cross-sectional study
- observations made at single point in time

- retrospective study

eg. study 3.

Case-control study
- select cases of disease
- select controls - don't have disease

- look at their "history of exposures" ← before they got disease

establish temporal relationship.

observational design
- prone to confounding

advantage - don't have to wait - prospective study

case-control - not definitive
- good in initial stages

ex . 2x2 table

		Cancer Y	Cancer N	
Smoking	Y	41	28	69
	N	19	32	51
		60	60	120

response total fixed by design
outcome

mechanically could consider

$\hat{p}_1 = 41/69 = .59$

proportion of smokers with cancer

$\hat{p}_2 = 19/61 = .37$

proportion of non-smokers with cancer.

prevalence of cancer is 50%
by design

$p_1 - p_2$ · risk difference

can't be estimated in a case · control study

- what can we estimate??
  - odds ratios] - logistic regression