Stat 145, Fri 3-Sep-2021 -- Fri 3-Sep-2021
Biostatistics
Spring 2021

Friday, September 3rd 2021

Wk 1, Fr
Topic:: Associations between variables
Read:: Lock5 1.3

Warmup questions:

1. If bias can be unintentonal, how best to eradicate it?
2. If you record sex as 1 for female, 0 for male, is it now quantitative?
3. Is there ever a left-to-right progression for categorical values?

Association — what many studies are meant to discover

- discussed between variables: generally have two
- means knowledge of a subject's value for one variable makes you somewhat better informed about the value of other variable.
  Ans: yes or no
- explanatory variable and response variable
  height and weight
  sex and length of life
  number of bars in a community and the number of churches there
- not same as causality

Types of studies

- experiment vs. observational studies
- features of experiments
  Required:
    factor(s) and response
      so, at least 2 variables
      "factor" is synonym for "explanatory variable"
        1-, 2-, multi-factor experiments
        values of a factor are called its "levels"
        a set combination of levels for all factors is a "treatment"
    levels of factors are imposed, not merely observed

  Desirable:
    random assignment to treatments
      should even out instances of other variables across groups

terms
factor = explanatory var.
levels: the values of a factor
treatment

```
         --> thus, can establishing causation
        observational studies can establish association
          in observational studies, association may be due to lurking vars

     control/placebo group: a treatment will all factors set to zero
      blind study: those in control group do not know it (anti-bias)
      double-blind study: researchers also unaware (anti-bias)
      repetition:


relationships between variables:
 ssurv <- read.csv("http://scofield.site/teaching/data/csv/ssurv.csv")

 tally:
   tally(~selfhandedness | sex, data=ssurv)
   tally(~selfhandedness | sex, data=filter(ssurv, selfhandedness!=""))

   tally(~selfhandedness | dadhandedness, data=filter(ssurv, dadhandedness!=""))

 addmargins():
   addmargins(tally(~selfhandedness | dadhandedness, data=ssurv))     or
   tally(~selfhandedness | dadhandedness, data=ssurv) %>% addmargins()

 gf_bar()
   gf_bar(~selfhandedness|dadhandedness, data=filter(ssurv, dadhandedness!=""))

 gf_histogram()
   gf_histogram(~speedtickets | oncampus, data=ssurv)
   gf_histogram(~speedtickets | oncampus, data=filter(ssurv, oncampus!=""))

Associations between variables
 - context: bivariate data
    from ssurv.csv
      sex and handedness




      speeding tickets vs. off-campus
```

speeding tickets vs. number of cds

Features of experiments, 2nd look: what they are aimed at doing

- reduce bias
  1. control group
     group "identical" in all aspects to treatment group
  2. randomization
     how we make the groups "identical"
  3. blinding
     placebo effect
     single vs. double
- reduce sampling error
  1. replication
     larger sample size---can pinpoint effect more precisely
  2. balance
     sizes of treatment groups roughly equal
  3. blocking
     matched pairs: the blocks are the pairs
     agricultural experiments

See also

http://www.zoology.ubc.ca/~whitlock/bio300/overheads/overheads14.pdf