Distribution — values a variable takes and how often

Stat 145, Mon 8-Feb-2021 -- Mon 8-Feb-2021
Biostatistics
Spring 2021

Say two variables are associated
if knowing the value of one
for a case makes you somewhat
better-informed about the value of
the other for that case.

Monday, February 8th 2021

Wk 2, Mo
Topic:: Associations between variables
Read:: Lock5 1.3

Associations between variables
 - context: bivariate data
    from ssurv.csv
       sex and handedness   both are categorical: take gender/sex as explanatory
         tally (~ selfhandedness | gender, data = ss)   contingency (two-way) table
           ego
         gf_bar (~ selfhandedness | gender, data = ss)   side-by-side bar graphs

       speeding tickets vs. off-campus   oncampus is categorical, take it as explanatory
         tally (~ speedtickets | oncampus, data = ssurv)
         gf_histogram (~ speedtickets | oncampus, data = ssurv)   side-by-side histograms

       speeding tickets vs. number of cds   both variables are quantitative
         gf_plot (speedtickets ~ cds, data = ssurv)
              creates a scatterplot
              uses cds as explanatory variable (horizontal axis)

 - explanatory and response
 - meaning: an association between 2 vars is what you have if
     knowledge of the value of a case's explanatory var. makes
     you better informed (even ever so slightly) about that case's
     response ~~variable~~ value
 - examples: Could pose as a research question whether an association exists between
1. sex and pulse rate
2. attend church at last twice a month and political affiliation

3. father's dominate hand and own dominate hand
4. number of bars in a community and number of churches

    Activity: Etherpad, name s145-08feb2021-gX
     identify an explanatory variable and a response
     indicate a summary/plot/table/means for investigating
 - don't confuse with causation
    may be a lurking/confounding variable (common cause?)
    types of studies: observational vs. experiment
      experiment involves manipulating the explanatory variable
      gold standard: controlled randomized experiment

Features of experiments, and what they are aimed at doing
 - reduce bias
    1. control group
      group "identical" in all aspects to treatment group
      "patients" seek treatment when feeling bad, improvement inevitable(?)
    2. randomization
      how we make the groups "identical"
      assignment in some nonrandom manner can be intended to reveal effect
    3. blinding
      placebo effect
      single vs. double
      unblinded studies typically find larger effects than blinded ones
 - reduce sampling error
    1. replication
      larger sample size---can pinpoint effect more precisely
    2. balance
      increases precision by more carefully pinpointing effect for each group
    3. blocking
      matched pairs: the blocks are the pairs
      agricultural experiments
      illustration: see final page of document at

Read Lock 1.3
See also
  http://www.zoology.ubc.ca/~whitlock/bio300/overheads/overheads14.pdf