Stat 145, Fri 1-Oct-2021 -- Fri 1-Oct-2021
Biostatistics
Spring 2021

Friday, October 1st 2021

Wk 5, Fr

Topic:: Bootstrapping works
Topic:: Bootstrapping on differences
Topic:: Bootstrap coverage rates
Topic:: Bootstrap for slope, correlation
Topic:: Bootstrap percentile CIs
Read:: Lock5 3.4
HW:: PS06 due Tues.

Bootstrapping CI
- what parameters aiming for? what are the corresponding bootstrap stats?

params	bootstrap stats	bootstrap samples drawn
μ	$\bar{x}$	with replacement, from slips, one per x_i (sampled quantitative resp.)
p	$\hat{p}$	with replacement, from slips, one per sampled categorical response
$\mu_1 - \mu_2$	$\bar{x}_1 - \bar{x}_2$	with repl., from two sets of slips containing quant. values from 2 groups
$p_1 - p_2$	$\hat{p}_1 - \hat{p}_2$	with repl., from two sets of slips containing categ. values from 2 groups
ρ	r	with replacement, from slips, one per sampled (x_i, y_i)
β_1	b_1	with replacement, from slips, one per sampled (x_i, y_i)

- StatKey software: Note the options of the "bootstrap confidence interval" column

- Percentile confidence intervals: See below

- "bag" metaphor for bootstrapping
  case: when interested in mu1 - mu2
  two bags, one contains n_1 slips of sampled values from group 1
  the other contains n_2 slips of sampled values from group 2
  draw, with replacement, n_1 slips from Bag 1, use to calculate $\bar{x}_1$
  draw, with replacement, n_2 slips from Bag 2, use to calculate $\bar{x}_2$
  } Then subtract $\bar{x}_1 - \bar{x}_2$

case: slope or correlation

One bag containing n slips

data is matched pairs, each case contributes pair (x_i, y_i)

every slip in bag has one (x_i, y_i) - pair from sample

Draw n times w/ replacement getting bootstrap sample of n (x_i, y_i) - pairs

most likely there are repeat pairs in bootstrap sample

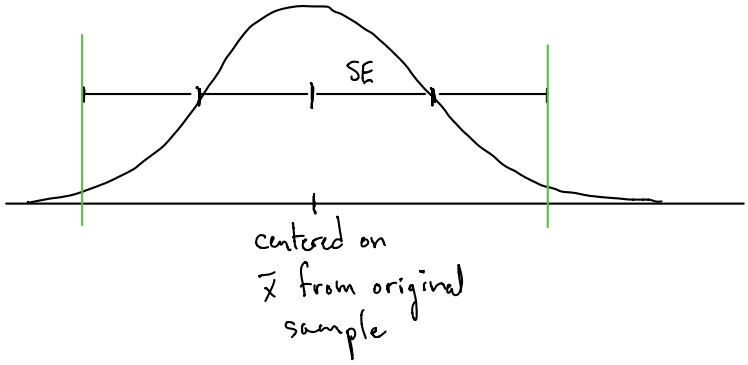
use to calculate r or b_1 (bootstrap statistic)

Bootstrap percentile confidence intervals

Idea: 95% confidence interval with a truly normal bootstrap distribution when done using the centered interval approach

(pt. estimate) $\pm$ 2(SE)

would result in a lower bound at the 2.5-percentile, and an upper bound at the 97.5-percentile of the bootstrap distribution.



Alternatively, use tools of software to compute directly

2.5 and 97.5-percentiles – for 95% confidence

10 and 90-percentiles – for 80% confidence

5 and 95-percentiles – for 90% confidence

etc.