Stat 145, Wed 3-Mar-2021 -- Wed 3-Mar-2021
Biostatistics
Spring 2021

Wednesday, March 3rd 2021

Wk 5, We
Topic:: Bootstrap coverage rates
Topic:: Bootstrap for slope, correlation
Topic:: Bootstrap percentile CIs

Groups discuss answers, and why correct, from socrative quiz

For test, want students to be able describe physical process of bootstrapping
 - cases for
    single mean
    single proportion
 - in full detail, including
    the setup: sample, well-mixed slips in a bag, what is on a slip
    how bootstrap sample is drawn
    what bootstrap sample is used for
    how you get a full bootstrap distribution

Bootstrapping for slope/correlation
 - automatically means two quantitative variables are involved
 - parameters
    rho = true correlation between variables
    beta = true slope of regression line between variables

    can compute these parameters only when you have full population
 - illustration using MLB 18 data: RBI and H (hits) variables
    consider mlb18 to contain full population data
    commands for parameters

```
cor(RBI ~ H, data=mlb18)        # gives rho
lm(RBI ~ H, data=mlb18)         # can read beta
```

    commands when using samples of size n=40

```
mySample <- sample(mlb18, size=40)
cor(RBI ~ H, data=mySample)        # gives r, sample stat
```

```
    lm(RBI ~ H, data=mySample)$coefficients[2]   # can read b, sample stat

  sample correlation/slope have sampling distributions
    manyCors <- cor(RBI ~ H, data=sample(mlb18, size=40))
        not quite as symmetric?

- But, we don't usually have a full population available
  bootstrap: do in StatKey
  description with bags
  bootstrap percentile confidence interval
    95% of values lie between .025-quantile and .975 quantile
    90% of values lie between .05-quantile and .95 quantile
```

Bivariate quantitative data

- Scatter plot
- looks linear?

  calculate Correlation

  find least-squares regression line

	Correlation	slope	intercept
sample	r	b	a
population	ρ	β	α