Stat 145, Wed 22-Sep-2021 -- Wed 22-Sep-2021
Biostatistics
Spring 2021

Wednesday, September 22nd 2021

Due:: Quiz Ch. 2 ends at 10 pm

Wednesday, September 22nd 2021

Wk 4, We
Topic:: Confidence intervals
Read:: Lock5 3.2

From activity yesterday:

- For any fixed quantitative population variable, there is a different sampling distribution for the sample mean x-bar for each fixed sample size n. As n increases, these sampling distributions change predictably ...
  - they shrink in their spread (SE $\downarrow$)
  - they look more symmetric bell-shaped (if they didn't already)
  - mean of sampling dist. always matches μ (mean of population).
    ($\Rightarrow$ $\bar{x}$ is unbiased estimator of μ)
- As n increases, the standard error of x-bar ...

  $SE_{\bar{x}}$ shrinks

  i.i.d. = independent and identically distributed

- When the sample size n corresponds to a relatively large portion of the overall population size, there can be a noticeable, if not always overly large, difference between sampling distributions for x-bar obtained via SRS samples and those obtained as i.i.d. samples. The one with the larger standard error is ...

  the sampling dist. when taking i.i.d samples (w/replacement)

```
Sampling distributions for
 - simulation requires repeatedly sampling from full population
    sampling method: SRS vs. i.i.d.


 - means
    centered on the population mean mu
    become increasingly normal (they may be already) as n grows
    applets
today  →  https://shiny.calvin.edu:3838/scofield/cltMeans/
      http://www.lock5stat.com/StatKey/index.html
    spread (SE) shrinks as n grows


 - proportions
    binary categorical variable
      proportion arises from looking for one of two values as a "success"
    applets
today  →  https://www.rossmanchance.com/applets/2021/oneprop/OneProp.htm?candy=1
       →  https://shiny.calvin.edu:3838/scofield/cltProportions/
Monday →  http://www.lock5stat.com/StatKey/index.html
    become increasingly normal as n grows
    spread (SE) shrinks as n grows
See video at
http://scofield.site/courses/s143/videos/samplingDistsProportionsFirstLook.mp4

Normal distributions
 - Sampling distributions for sample means/proportions tend to look "normal"
    truer when sample size n is large
 - Normal distribution calculator from StatKey
 - 68-95-99.7% rule
    In particular,
      about 95% of values of p-hat lie within 2 standard deviations of p
        that is, inside  [p - 2 SE, p + 2 SE]
      about 95% of values of x-bar lie within 2 standard deviations of mu
        that is, inside  [mu - 2 SE, mu + 2 SE]
      call the amount added/subtracted the "margin of error" for 95% coverage

Q1: Use the Normal distribution calculator app from StatKey to
    (a) Plot a Normal distribution with mean 100 and std dev 20.
    (b) Find rel. freq of values from this distribution between 90 and 110
```

Maybe use as

warmup on Friday

What is the Z-score of 110?

(c) Find rel. freq of values from this distribution between 80 and 120
What is the Z-score of 80?

(d) Find rel. freq of values from this distribution between 60 and 140
What is the Z-score of 140?

(e) Find rel. freq of values from this distribution between 40 and 160
What is the Z-score of 40?

Q2: Say that a company fills its packages to an avg of 48 Kg with sd 2 Kg.

(a) Plot a Normal distribution with these parameters

(b) Find rel. freq of values from this distribution between 47 and 49
What is the Z-score of 49?

(c) Find rel. freq of values from this distribution between 46 and 50
What is the Z-score of 46?

(d) Find rel. freq of values from this distribution between 44 and 52
What is the Z-score of 52?

(e) Find rel. freq of values from this distribution between 42 and 54
What is the Z-score of 42?

Idea of a 95% CI (Centered-interval construction method)

- get an estimate for population parameter
  in case of mu (quant var), use x-bar
  in case of p (binary categorical var), use p-hat
- get an estimate for margin of error (ME) for 95% coverage
  most likely this involves estimating SE and doubling it
- construct centered interval

Q3: SE for 95% coverage is approximately 2*ME.
How should you get SE for
99.7% coverage?
68% coverage?
100% coverage?
90% coverage?