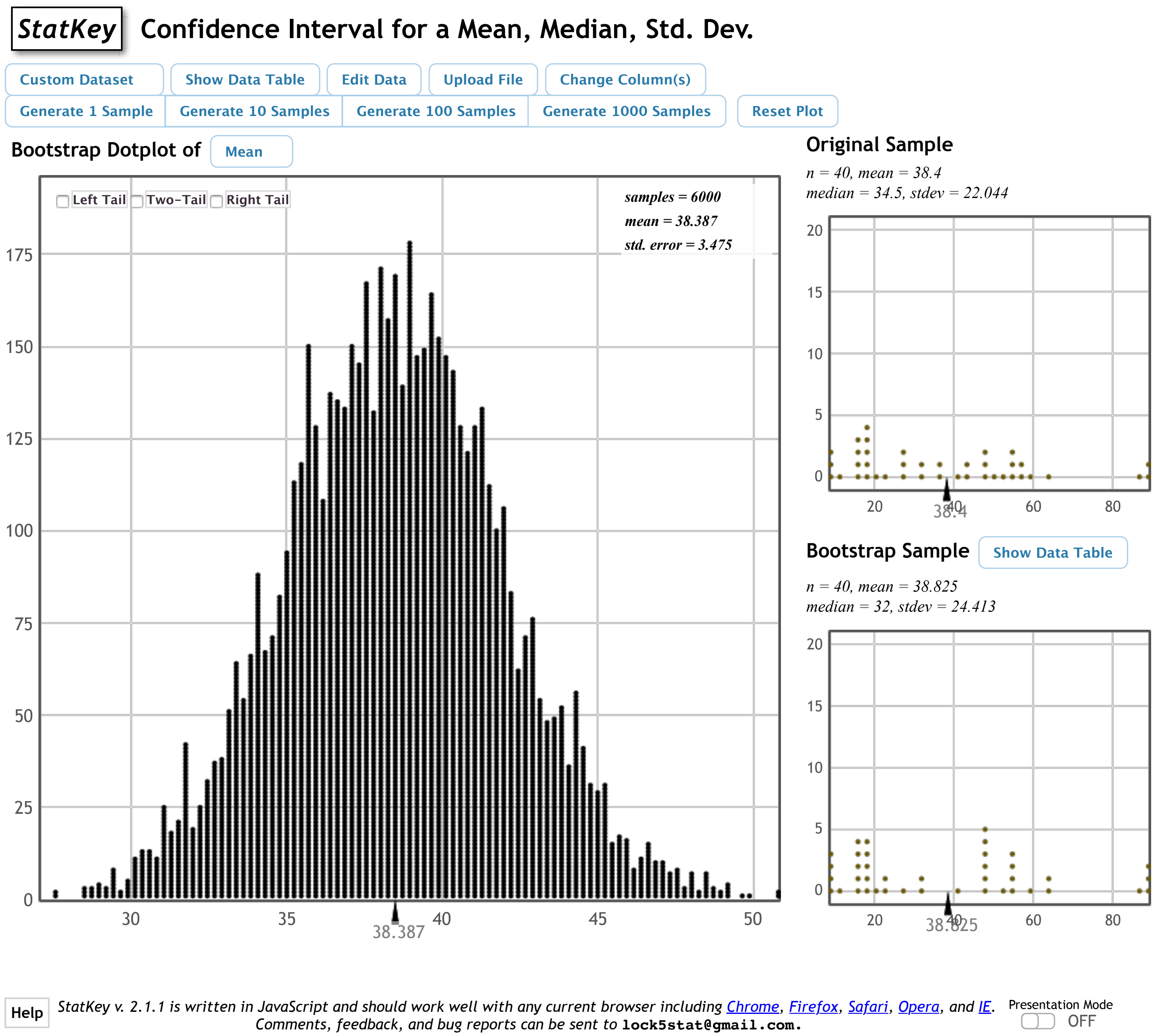
Preview (not done live in class today)



Pictured above is a bootstrap distribution where each dot represents the mean of a bootstrap sample. The original sample appears in the top right panel, is of a quantitative variable with unknown population mean μ; the original sample mean $\bar{x} = 38.4$. 6000 bootstrap samples, all of the same size $n = 40$, were drawn and used to calculate 6000 $\bar{x}$-values, resulting in the dot plot. The std deviation of this bootstrap distribution, also called std. error, appears as $SE_{\bar{x}} = 3.475$.

Parameter: the thing we often want to know

1. p = proportion
   = relative freq. of certain value of categorical variable in population

   point estimate: $\hat{p}$ = proportion in random sample

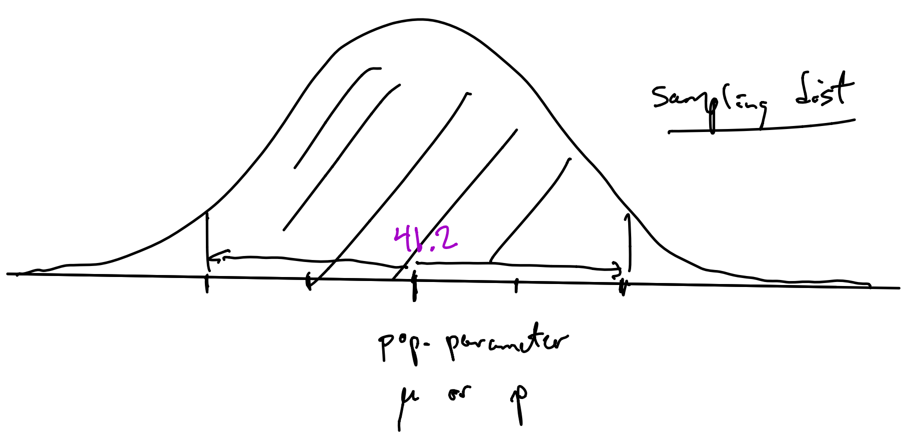
   Know: sampling dist. of $\hat{p}$ approx. normal if

$$\left.\begin{array}{l} np \\ n(1-p) \end{array}\right\} \text{both at least 10}$$

2. μ = mean
   = avg. of values of quant. variable in a population

   point est. $\bar{x}$ = sample mean

   If $n \geq 30$, $\bar{x}$ will have approx. normal dist.
   (sometimes less than 30 works well enough)



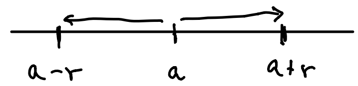
68-95-99.7% rule

2 std. deviations — more precisely, its 1.96 std. deviations in each direction

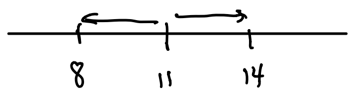
Building centered intervals

center : a

radius = "margin of error" r



Ex.) Centered interval w/ center = 11, radius = 3



$[8, 14]$

Ex.) Centered interval: center = 17, margin of error = 8.1

$[8.9, 25.1]$

↑ lower bound

upper bdry

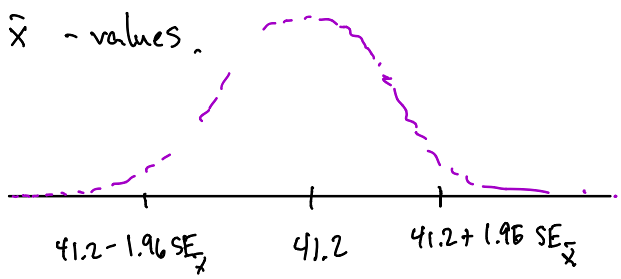
Ex.) Center: population parameter $\mu = 41.2$

Sampling dist. for $\bar{x}$ (some fixed sample size, say $n = 52$)

has std. deviation 3.8 $(SE_{\bar{x}} = 3.8)$

conclude normal-shaped dist.

Want a radius large enough (but no more than needed) to get a centered interval comprising 95% of $\bar{x}$-values.



$[33.75, 48.65]$

Putting these ideas to use:

Won't have population parameter (μ, p, etc. unknown)

Have sample estimate for that parameter: $\bar{x}$, $\hat{p}$, etc.

$$\boxed{(\text{point est.}) \pm \text{margin of error}}$$ Confidence Interval for param.

margin of error chosen to comprise certain pctg. of values

Case <u>95%</u>: margin of error $= (1.96)(SE)$

90%: " " " $= (1.645)(SE)$

99%: " " " $= (2.576)(SE)$

Ex.) Say want a 90% confidence interval for μ

having $\bar{x} = 119.5$ from sample of size $n = 80$

and $SE_{\bar{x}} = 2.3$

Ans. lower bound $= 119.5 - (\overset{z^*}{1.645})(2.3) = 115.7$

upper bound $= 119.5 + (1.645)(2.3) = 123.3$

or $[115.7, 123.3]$.

Lingering question: How to approximate SE?

Answer: (Only possible in recent/computing times)

<u>Bootstrapping</u>.

Want SE

Q: What is it?

A: Std. dev. of sampling dist.

Q: How do you get a sampling dist? (fixed n)

Schematic for simulating a sampling distribution of a statistic

Population
unknown parameter
(μ, p, etc.)

sample (size n)

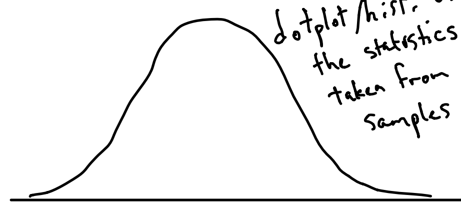
sample 1 $\longrightarrow$ compute sample stat. ($\bar{x}$, $\hat{p}$, ...)

sample 2 $\longrightarrow$ compute " "

⋮

possible to imagine sampling this often, but impractical to do.

Repetitively (thousands of times) returning to population to draw new sample of size n

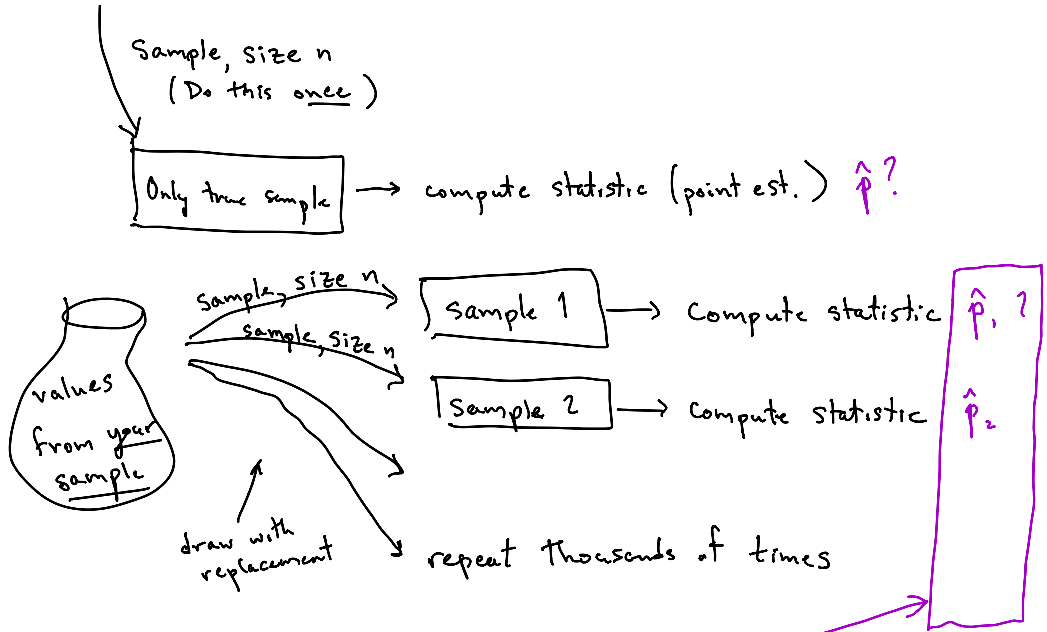


dotplot/hist. of the statistics taken from these samples

SE = std. dev. of sampling distribution

Bootstrapping

Population
unknown param.



Together, these sample stats (each known as a bootstrap statistic) a bootstrap distribution

Two important facts:

1. Sampling distribution centered at population parameter but bootstrap distribution centered at sample statistic of original sample

2. But, std. deviation of bootstrap dist. $\approx$ std. dev. of sampling dist., so the std. dev. of the bootstrap

distribution can be used to approximate SE.