Stat 88 lec 29

warmup 2:00-2:10

exercise 9.5.6

6. "One pound" bags of coffee produced by a local company actually contain a random amount of coffee. The amounts of coffee in the bags are i.i.d. with unknown mean μ.

In 100 bags the average amount of coffee is 16.3 ounces and the SD is 0.7 ounces. For comparison, an ounce is approximately the weight of five quarters.

Construct an approximate 95% confidence interval for the underlying mean μ. Do the data indicate that $\mu = 16$ ounces?

X = weight in pounds of coffee bag in ounces

$X_1, \dots, X_{100}$ iid mean μ, SD σ

$\bar{X} = 16.3$, $\sigma = .7$, $n = 100$

95% CI is $\bar{X} \pm 2\frac{\sigma}{\sqrt{n}} = 16.3 \pm 2\left(\frac{.7}{10}\right) = 16.3 \pm .14$

$= \boxed{(16.16, 16.44)}$

Data indicates you would reject the null $\mu = 16$, at level .05, because 16 isn't in your 95% CI.

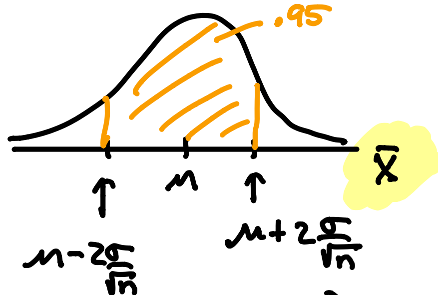
Last time

Sec 9.3 Confidence Interval for μ

$X_1, \ldots, X_n$ iid with mean μ, SD σ

Assuming some H_0

Sampling distribution



$$P\Big(\underbrace{\mu - 2\tfrac{\sigma}{\sqrt{n}} < \bar{X} < \mu + 2\tfrac{\sigma}{\sqrt{n}}}_{\text{acceptance region for } H_0 \text{ at level .05}}\Big) = .95$$

$$\Leftrightarrow \quad P\Big(\underbrace{\bar{X} - 2\tfrac{\sigma}{\sqrt{n}} < \mu < \bar{X} + 2\tfrac{\sigma}{\sqrt{n}}}_{95\% \text{ CI for } \mu.}\Big) = .95$$

Notice that $\bar{X}$ is in the acceptance region of H_0, at level .05, exactly when μ is in your 95% CI.

So in warmup "Do the data indicate $H_0: \mu = 16$"

you said <u>no</u> because 16 wasn't in the 95% CI.

Today

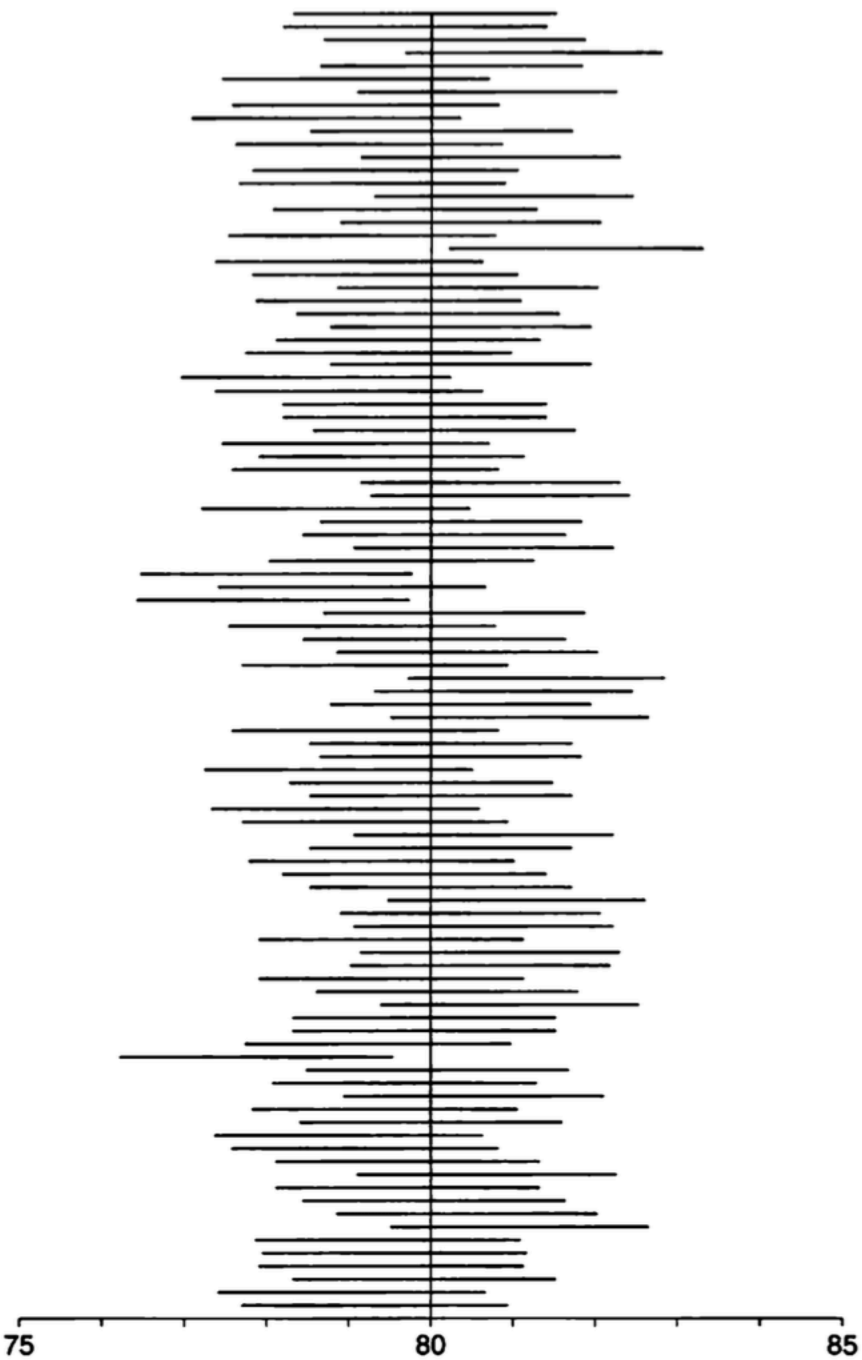
① Sec 9.4 interpretation of 95% CI of μ

② Sec 10.1 density

① Sec 9.4 interpretation of 95% CI of μ

ex Draw 100 marbles from a jar having 80% red marbles.
There is approximately a 95% chance your 95% CI contains 80%

Figure 1. Interpreting confidence intervals. The 95%-confidence interval is shown for 100 different samples. The interval changes from sample to sample. For about 95% of the samples, the interval covers the population percentage, marked by a vertical line.[8]

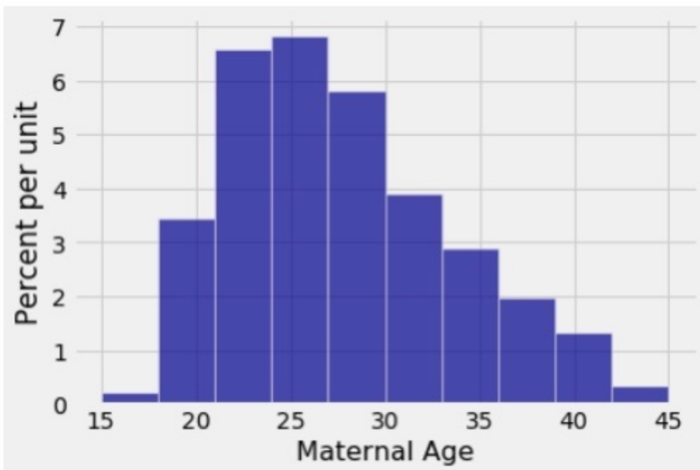


ex Suppose your 95% CI is (79, 82). Is there a 95% chance $\mu \in (79,82)$? — No, μ is a fixed number, 80%, and (79,82) is a fixed interval.

Comparison with bootstrap

The interpretation of CI is the same as in Data 8

ex Here is a distribution of 1174 maternal ages (years) from a random sample. $\bar{X} = 27.23$, $SD(X) = 5.8$



Find the approx. 95% CI of μ and interpret.

$$27.23 \pm 2\left(\frac{5.8}{\sqrt{1174}}\right)$$

$$= (26.89, 27.57)$$

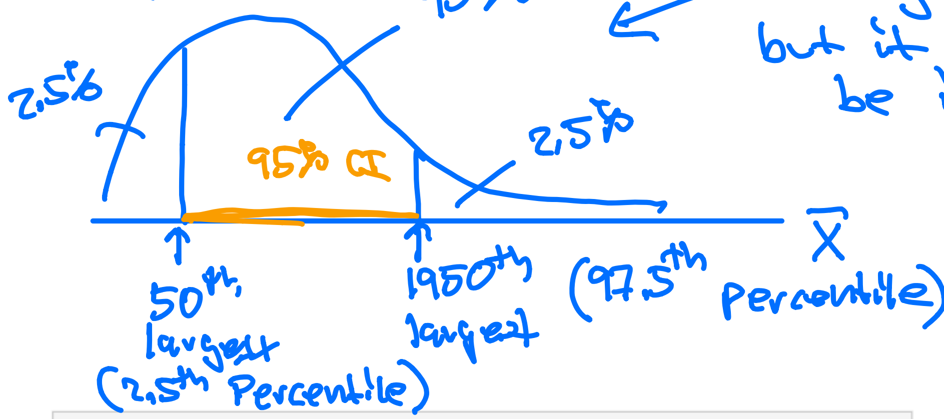
There is a 95% chance a CI contains μ, but my CI either does or doesn't contain μ.

This works because $\bar{x}$ is normally distributed by CLT. But if n is too small $\bar{X}$ may not be normal and we have to bootstrap your 95% CI.

How do you do this?

draw a large number (say 2000) samples of size 1174 with replacement from our original sample. Find the 50th and 1950th largest $\bar{x}$. This interval is a bootstrap 95% CI.

Picture



usually $\bar{X}$ is normal but it might not be if n is too small.

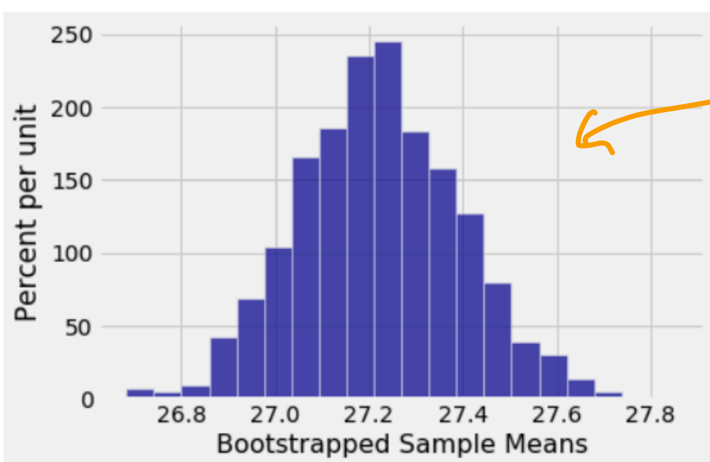
```
def one_resampled_mean():
    return np.average(births.sample().column('Maternal Age'))
```

We then called this function repeatedly to create an array of 2,000 bootstrap means:

```
means = make_array()

for i in np.arange(2000):
    means = np.append(means, one_resampled_mean())

Table().with_column('Bootstrapped Sample Means', means).hist(0, bins
```



approx normal by CLT.

Finally, we found the "middle 95%" of the bootstrapped means. That was our empirical bootstrap 95% confidence interval for the population mean.

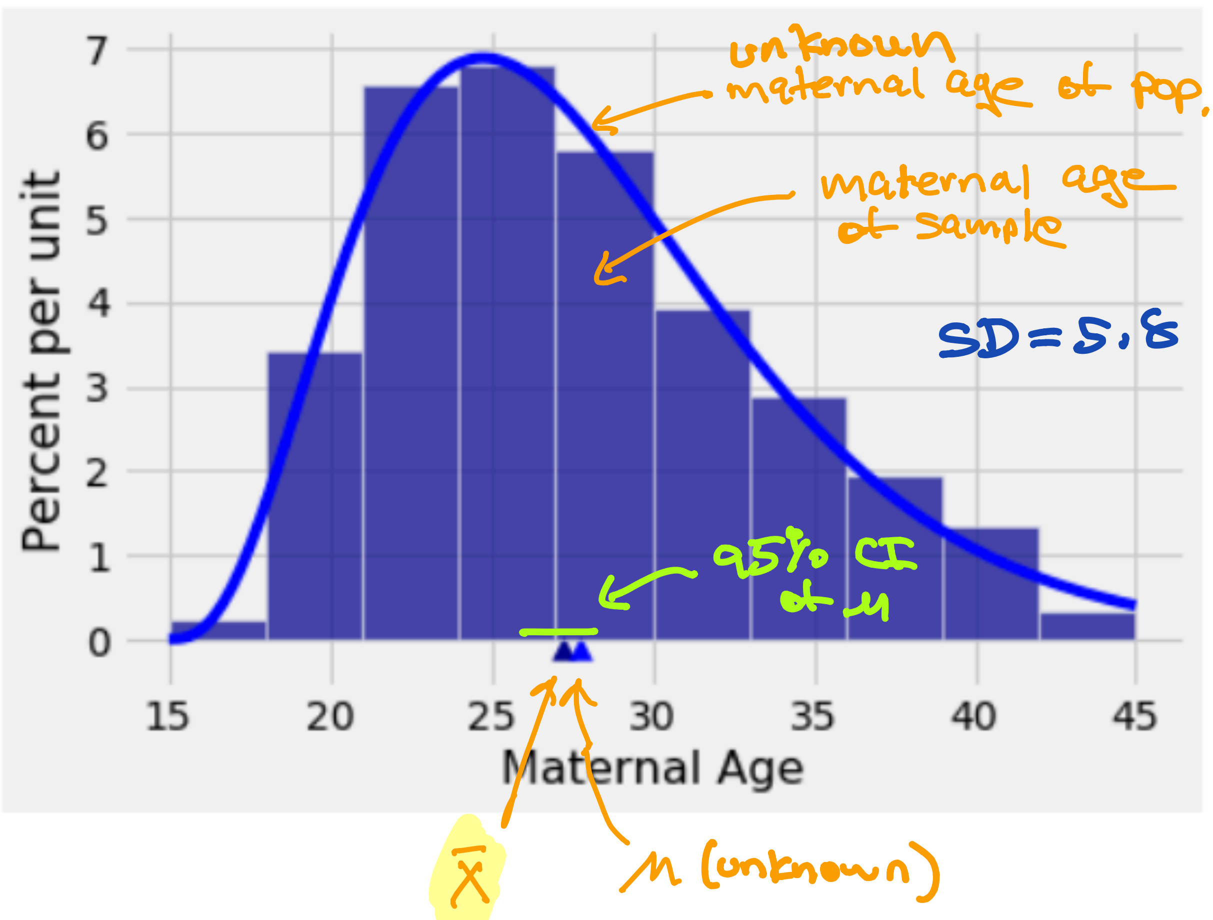
```
left = percentile(2.5, means)
right = percentile(97.5, means)
left, right
```

```
(26.89182282793867, 27.572402044293014)
```

close to (26.89, 27.57)

What the confidence interval measures

CI is an interval of estimates of μ



$\bar{X}$ is close to μ. On average it is $SD(\bar{X}) = \frac{\sigma}{\sqrt{n}}$ away from μ.

Is there a 95% chance that maternal ages are between

(26.89, 27.57)

No maternal ages are in a much wider range of values,

exercise 9.5.12

$CI = 68{,}000 \pm 2\frac{\sigma}{\sqrt{n}}$ $\quad \sigma = 40{,}000$ $\quad \sqrt{n} = 20$

$= \boxed{68{,}000 \pm 4000}$

12. A survey organization takes a simple random sample of 400 adults in a city. The annual incomes of the sampled people have an average of 68,000 dollars and an SD of 40,000 dollars.

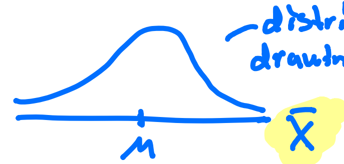
$\bar{X} = 68{,}000$, $\sigma = 40{,}000$

a) Fill in the blank with one of the words "sample" or "city".

The interval "68,000 dollars ± 4,000 dollars" is an approximate 95% confidence interval for the average annual income of adults in the city_____________.

b) Pick all of the correct options and justify your choices. More than one option may be correct.

The normal curve used in the construction of the confidence interval in Part **a** is the distribution of



- distribution of $\bar{X}$ drawing 400 from the city

(i) the incomes of the adults in the city

(ii) the incomes of the adults in the sample

(iii) the averages of all possible simple random samples of 400 adults from the city

c) True or false (explain):

68,000 ± 4000 is a 95% CI. — this is all you can say that is correct.

The incomes of approximately 95% of the adults in the city are in the range 68,000 dollars ± 4,000 dollars. — F

d) Fill in the blanks with the best choices you can make from the following set. You are welcome to use the same entry more than once.

- the average income of adults in the city
- the average income of adults in the sample
- 68,000 dollars
- 40,000 dollars
- 2,000 dollars

(400 adults) (their average income)

If you draw one adult at random from the city, that person's income has expectation equal to 68,000 and SD approximately equal to 40,000.

$\left(\frac{\sigma}{\sqrt{400}} = \frac{40,000}{20} = 2000\right)$