Confidence Intervals

Week 02 - Day 02

Common question during interviews

Sample Info

Population Info

Sample Info

Population Info + Uncertainty

Sample Mean = 7.2

Population Mean = ???

Point estimates vs. Ranges

Trump will get 53.2% of votes

Vs.

Trump will get between 51.2% and 55.2% of votes

Sample Mean = 7.2

Population Mean = 7.2 +/- 3.1

Sample Mean = 7.2

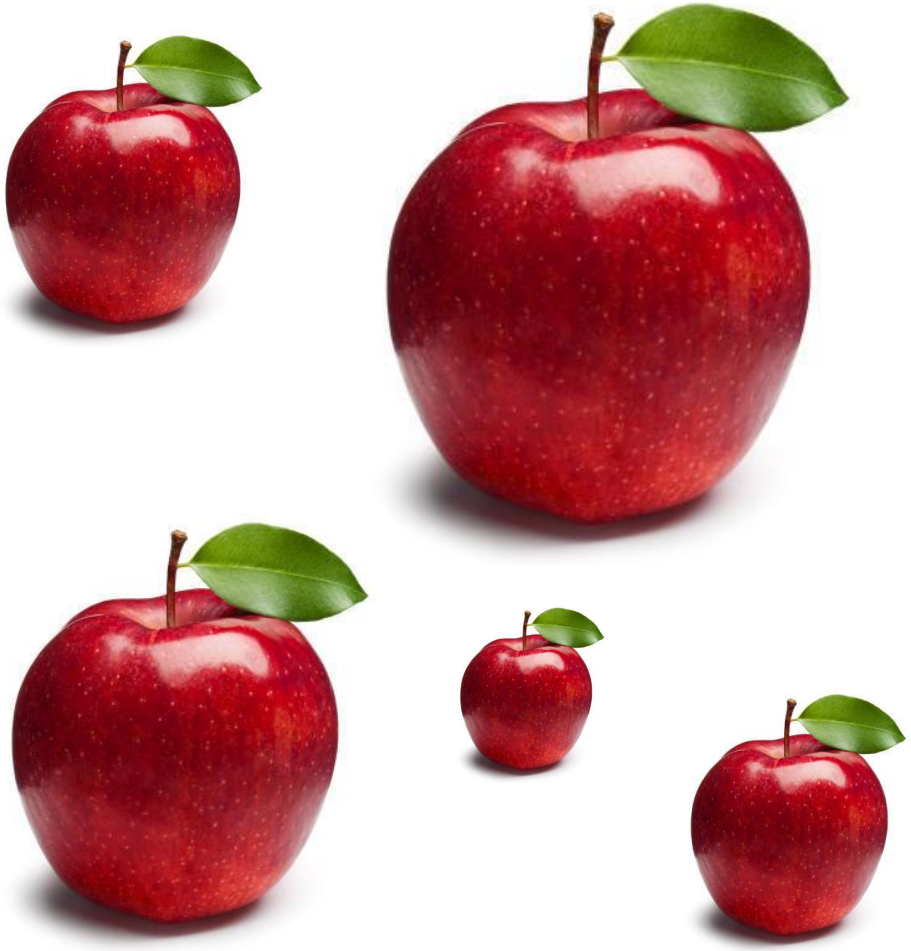
Population Mean = 7.2 +/- 3.1

What does it depend on??

Let’s talk about apples

Sample Mean = x

Population Mean = ???

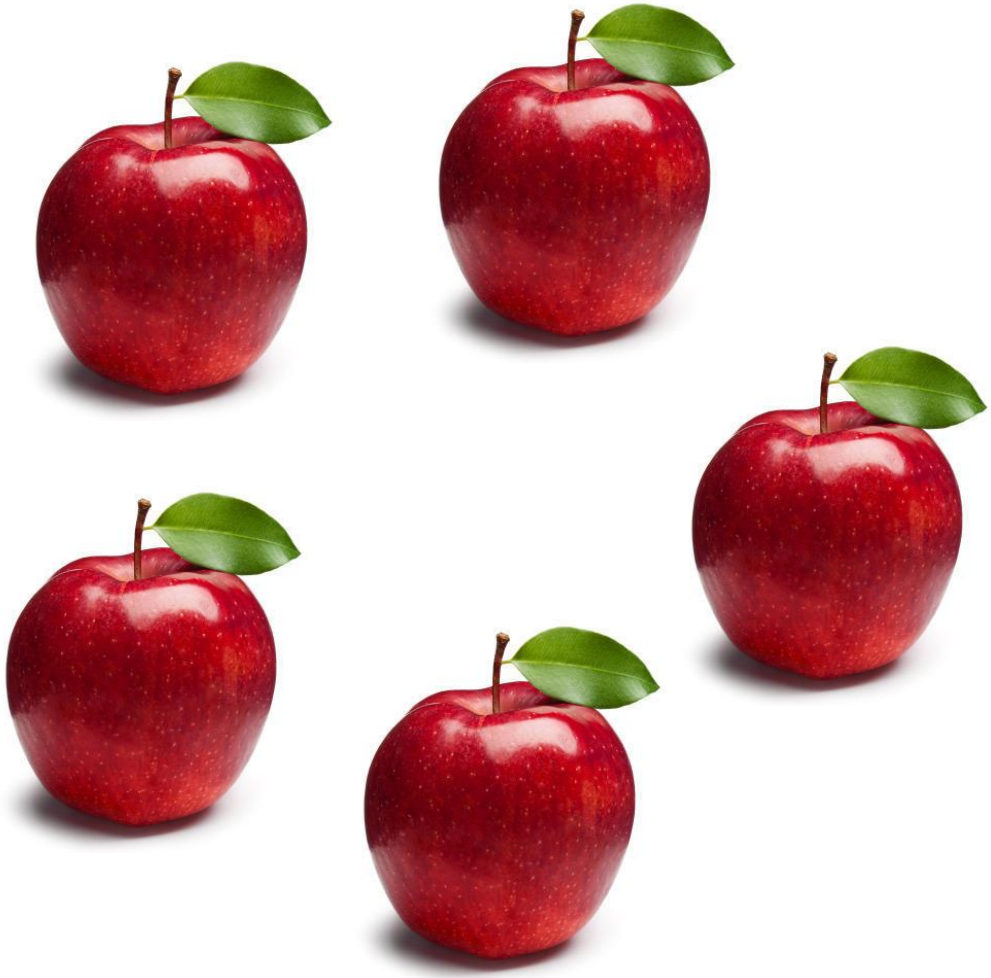


Population mean

5.5cm +/- 0.3

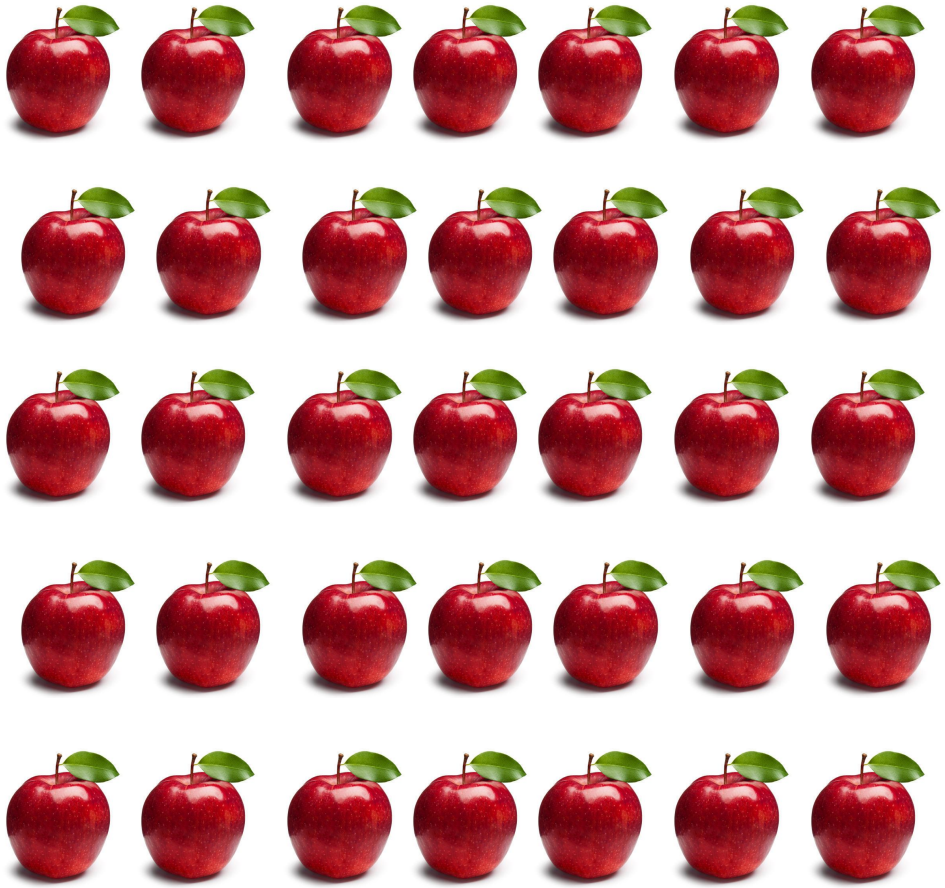
5.5cm +/- 1.5

???



Population mean

5.5cm +/- ???



Population mean

5.5cm +/- ???

Sample size + Sample STD influence the uncertainty!

Confidence Intervals

CI = how to build uncertainty

$$\mathrm{CI} = \bar{x} \pm z_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$$

$$\mathrm{CI} = \bar{x} \pm z_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$$

Sample mean

Sample std

$$\mathrm{CI} = \bar{x} \pm z_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$$

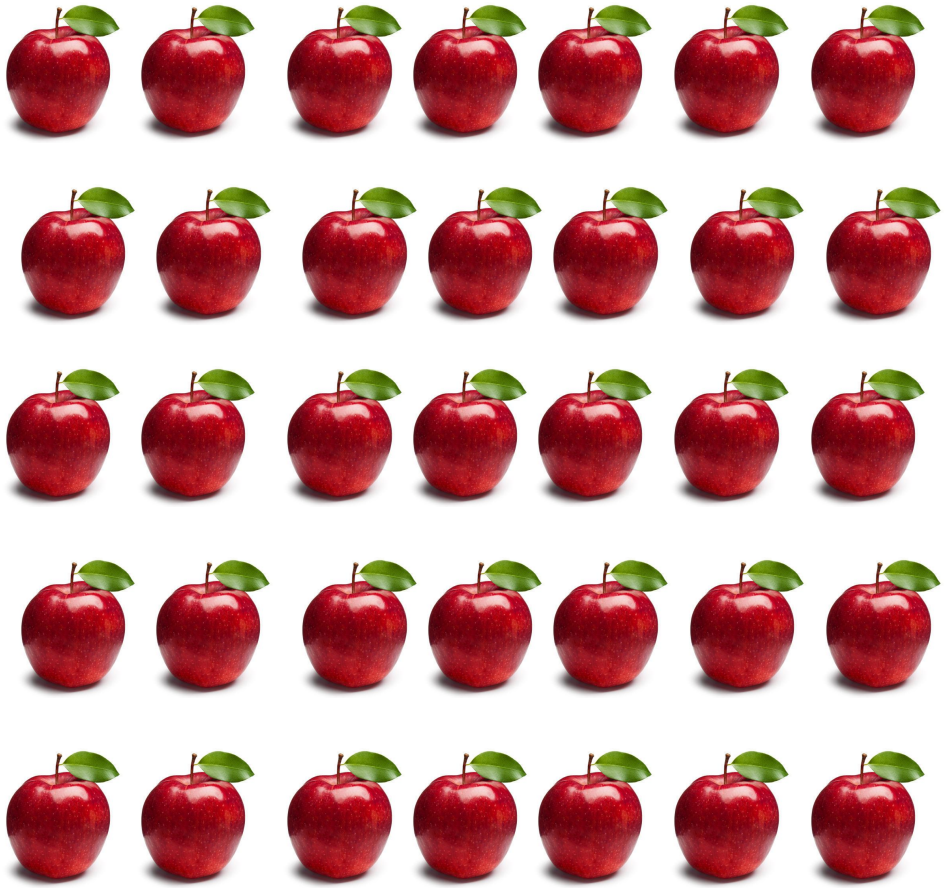
Sample mean

Sample std

$$\mathrm{CI} = \bar{x} \pm z_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$$

Sample mean

Sample size



Do you remember the apples?

High STD ➔ High uncertainty

$$\mathrm{CI} = \bar{x} \pm z_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$$

$$\mathrm{CI} = \bar{x} \pm z_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$$

Small sample ➜ High uncertainty

99% vs. 95% vs. 80%

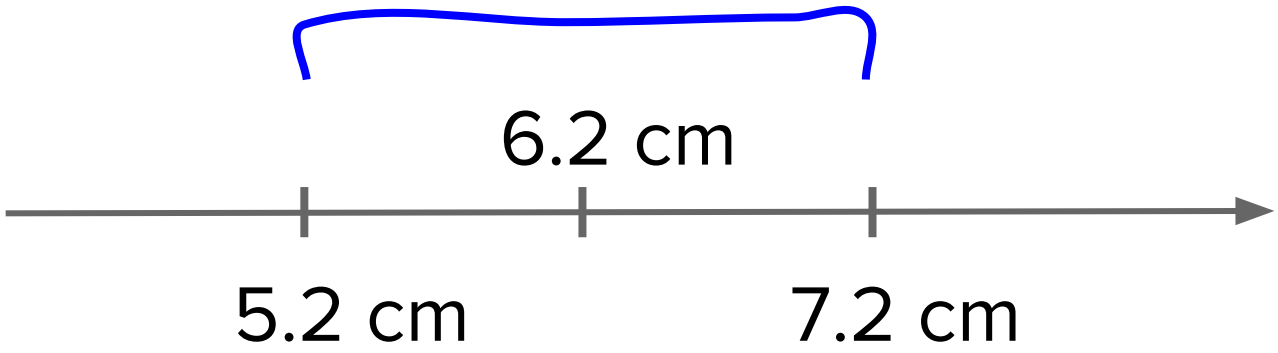
95% confidence interval

90% confidence interval

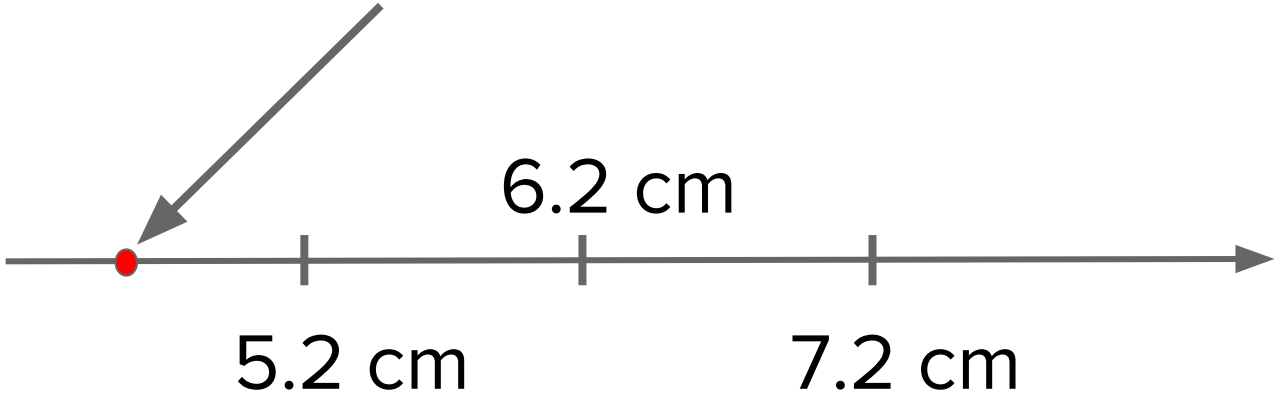
80% confidence interval

Uncertainty = things can go wrong

My nice confidence interval



The mean of the population can be here

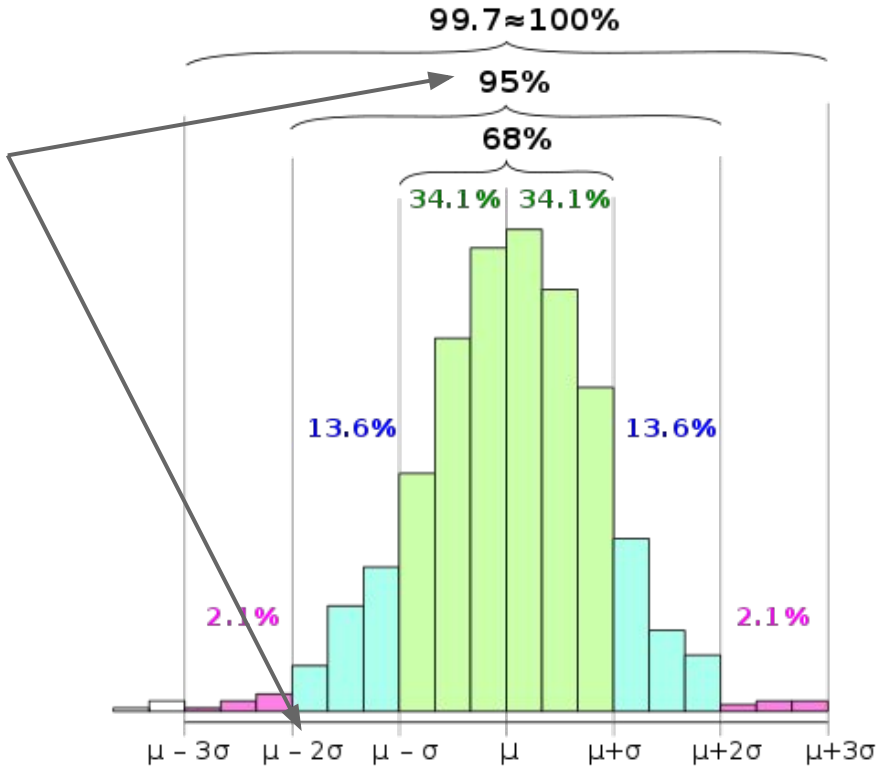


We want to be “95% sure the interval will contain the population mean”!

This is to control 95% vs 99% vs 80%

$$\mathrm{CI} = \bar{x} \pm z_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$$

Z = 1.96 for 95%



This is connected to the central limit theorem!

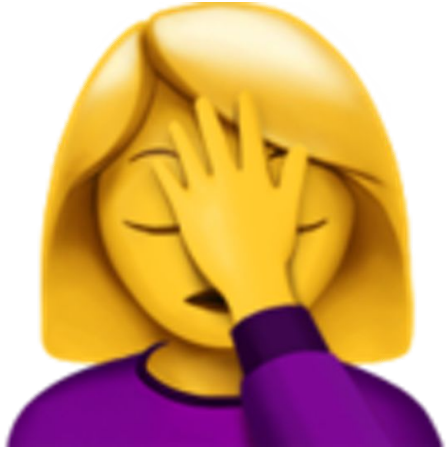
Standard error

$$\mathrm{CI} = \bar{x} \pm z_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$$

Why don’t we always use 99% instead of 95% or 80%?

I’m 99% confident your next salary

will be between $0 and $2M



I’m 80% confident your next salary will be between $40K and $70K

Interpretation

Common error during interviews

“There is a 95% probability that the real population mean is inside the range”

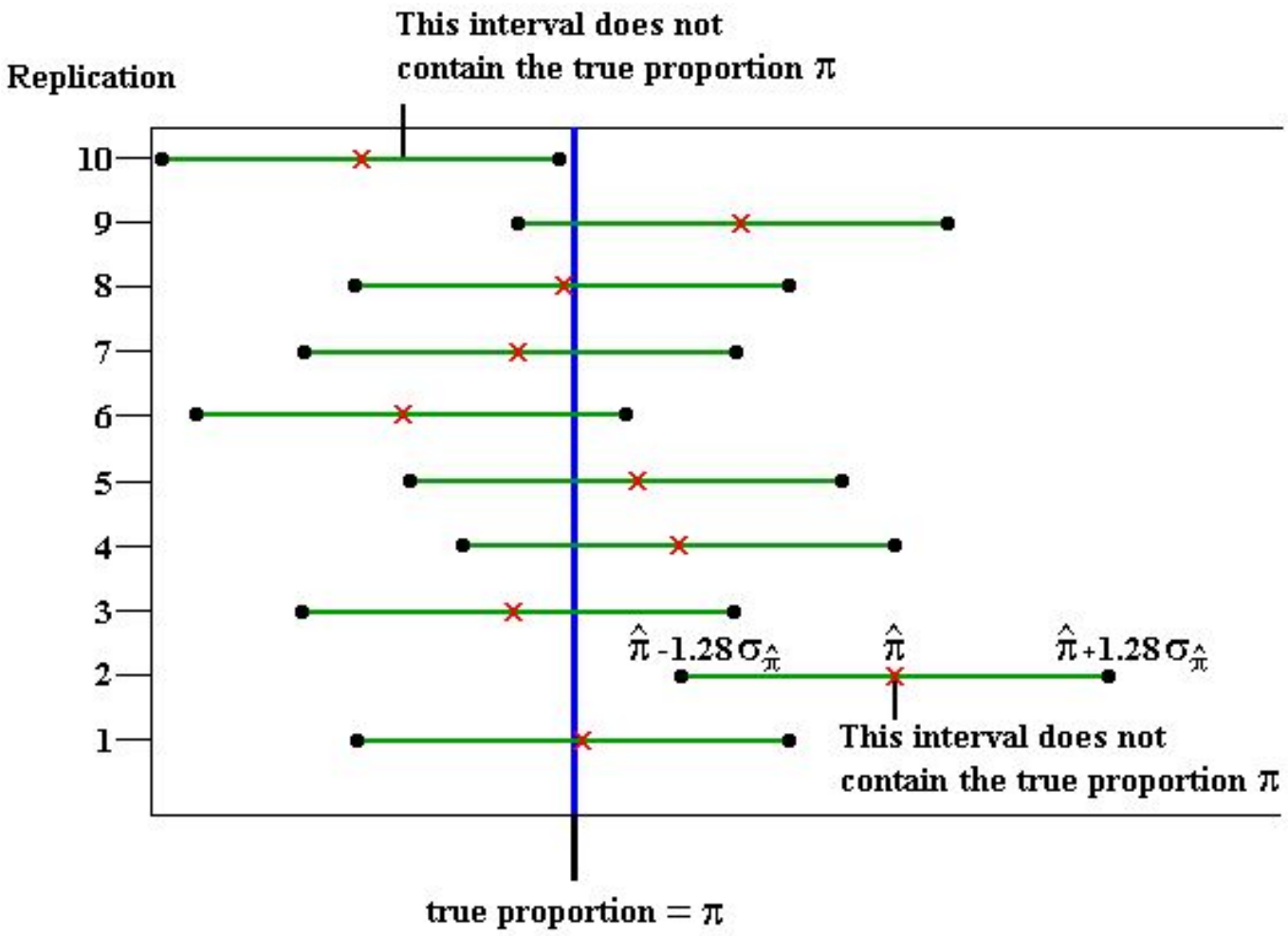
wrong!

~~There is a 95% probability that the real population mean is inside the range~~

Other similar samples

Other intervals

95% of the intervals will contains the population mean


This interval does not
contain the true proportion π
Replication
10
9
8
7
6
5
4
3
2
1
$\hat{\pi}$ -1.28 $\sigma_{\hat{\pi}}$
$\hat{\pi}$
$\hat{\pi}$ +1.28 $\sigma_{\hat{\pi}}$
This interval does not
contain the true proportion π
true proportion = π

Summary

- Sample -> population
- Uncertainty ~ std, sample size
- CIs use normal distribution and CLT
- 99% vs 95% vs 80%
- Right interpretation