NRES 776 Lecture 5

Sampling distribution and central limit theorem

Sunny Tseng

Our schedule today

- Announcement (5 min)
- Concept of sampling distribution (20 min)
- Central limit theorem (20 min)
- Wrap up (5 min)

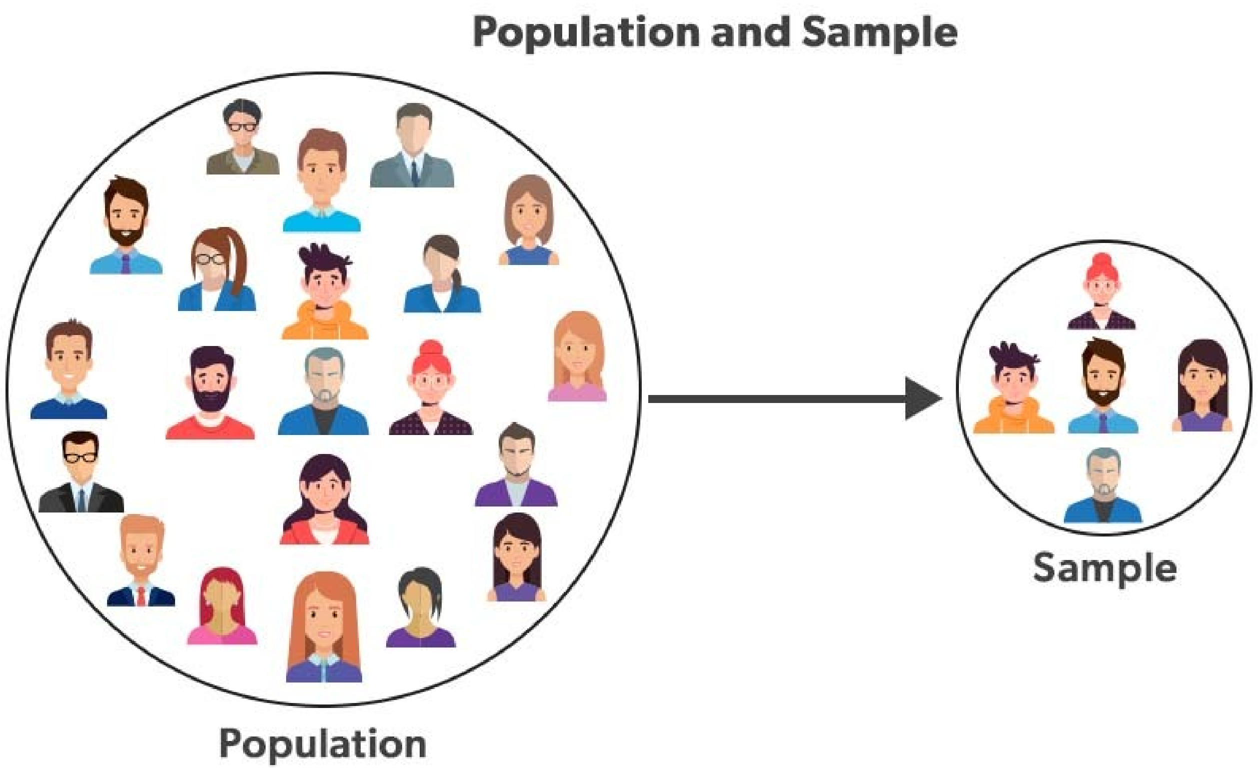
Announcement

- Let’s take this lecture easy…

Artwork by @allison_horst

Population and sample

If we are interested in the height of students in UNBC.



- How much can we trust the values we get from the sample?
- Is $\bar{x}$ close enough to μ?
- Is s_x some how represents σ?

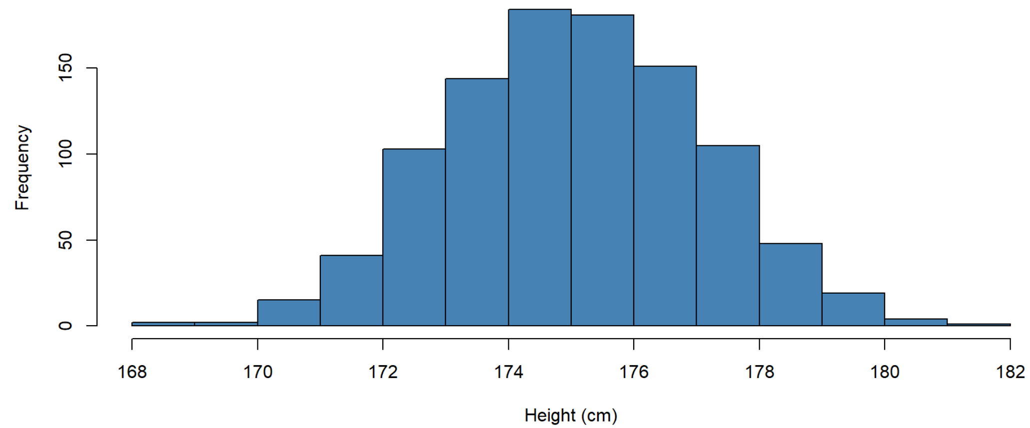
Mathematicians did a fun experiment

To answer whether we can trust $\bar{x}$

1. From a population, a group of people…
2. Randomly selected n people (sample 1), calculate $\bar{x}_1$
3. Put everyone back
4. Randomly selected n people (sample 2), calculate $\bar{x}_2$
5. Put everyone back
6. …(repeat above, say 1000 times)

Take a look at the $\bar{x}$s

- The distribution of $\bar{x}$ may look like this
- This is called **sampling distribution**

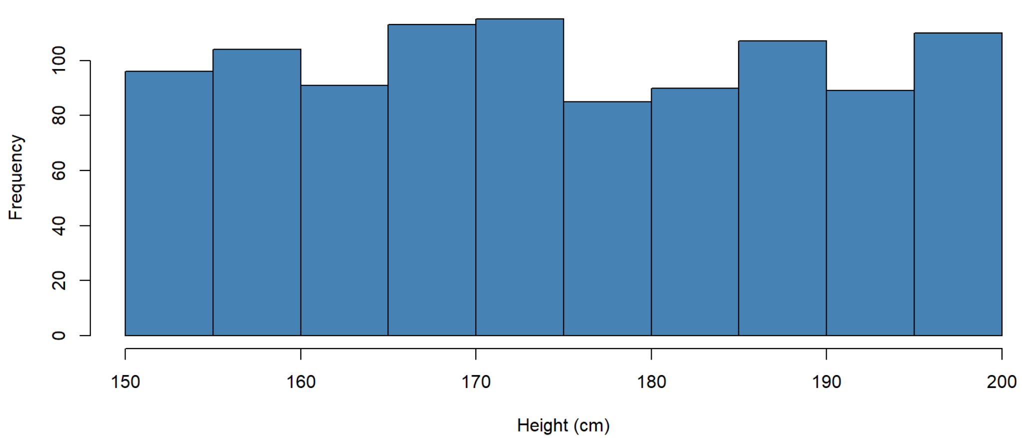


Sampling distribution

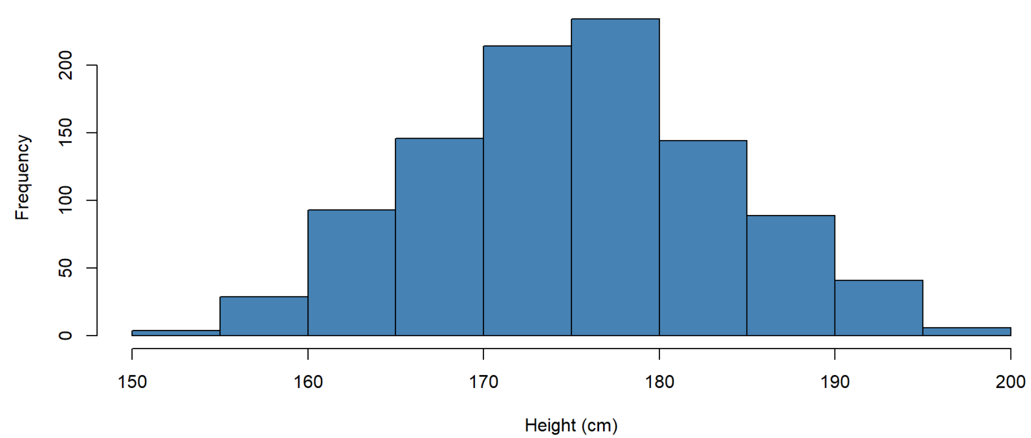
A probability distribution of a sample statistic that is obtained through repeated sampling of a specific population.

- What's special about this sampling distribution?
- Take a look at the n, our sample size

Population ($\mu = 175$)



Sampling distribution (n = 3)

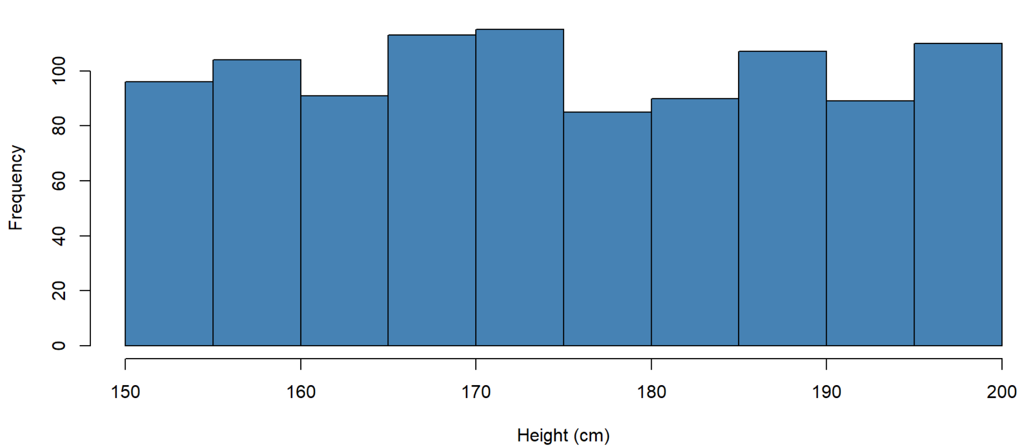


Sampling distribution

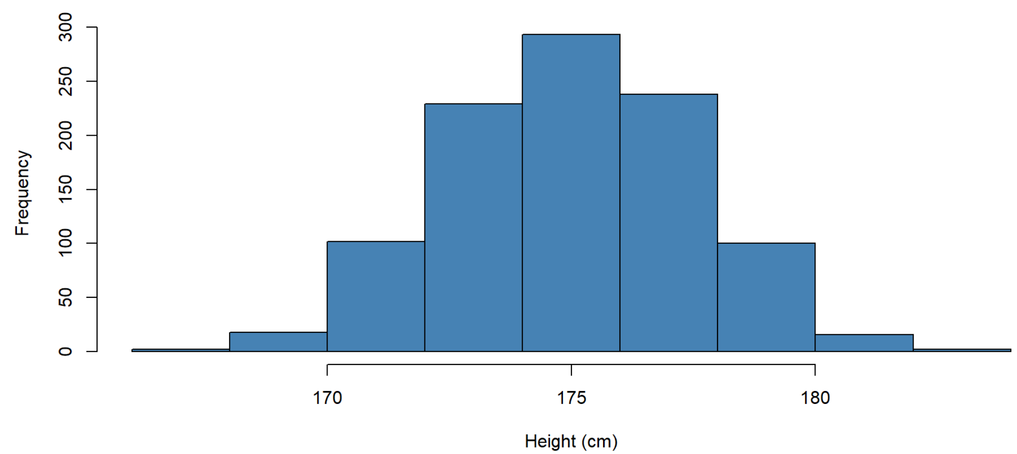
A probability distribution of a sample statistic that is obtained through repeated sampling of a specific population.

- What's special about this sampling distribution?
- Take a look at the n, our sample size

Population ($\mu = 175$)



Sampling distribution (n = 30)

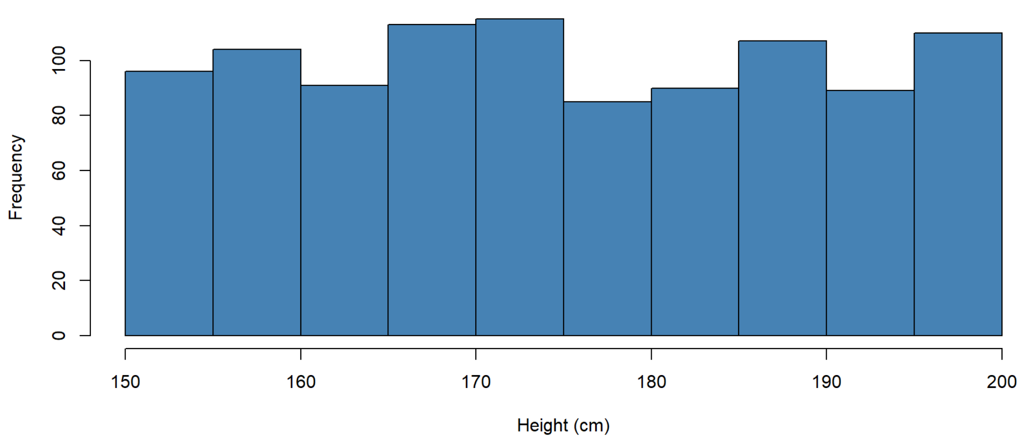


Sampling distribution

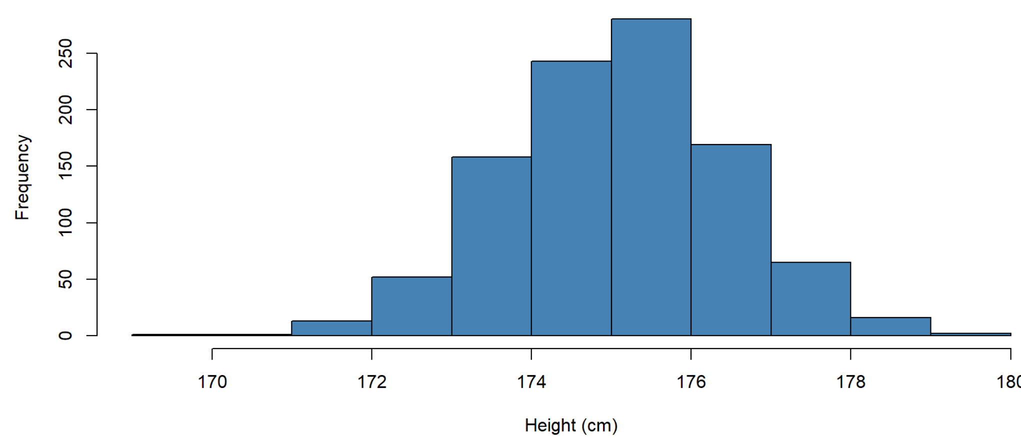
A probability distribution of a sample statistic that is obtained through repeated sampling of a specific population.

- What's special about this sampling distribution?
- Take a look at the n, our sample size

Population ($\mu = 175$)



Sampling distribution (n = 100)



Central limit theorem (CLT)

When $n \geq 30$, the sampling means are normally distributed with $mean(\bar{x}) = \mu$, and $sd(\bar{x}) = \frac{\sigma}{\sqrt{n}}$, no matter what distribution that population has.

- Central limit theorem is important and cool! Because…somehow, sampling distribution is almost always normally distributed, when n is large.

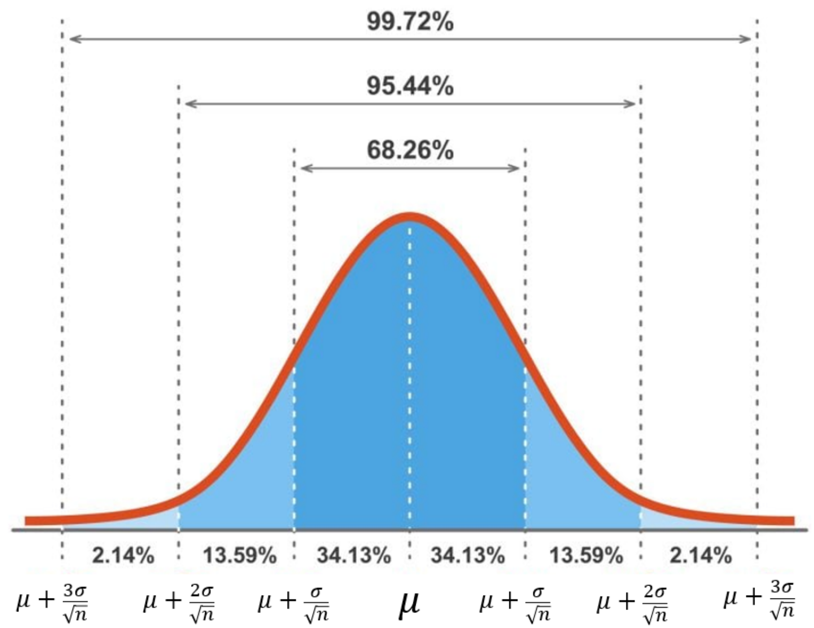
Standard error (SE)

Standard error is the standard deviation of the sampling distribution: $sd(\bar{x}) = \frac{\sigma}{\sqrt{n}}$

- SE describes standard deviation of sampling distribution
- SE means how wide spread is the 1000, or whatever many, sample means
- SE decreases when n increases; SE is large when σ is large

Now what?

- Based on CLT, imagine that you have a sampling distribution $N(\mu, \frac{\sigma^2}{n})$
- And you sample from population again, you get $\bar{x}_{new}$ as its mean
- $\bar{x}_{new}$ has 68% probability to locate within $\mu \pm \frac{\sigma}{\sqrt{n}}$, that is $\mu \pm \frac{s}{\sqrt{n}}$
- We can have a sense about population just by **ONE sample**!



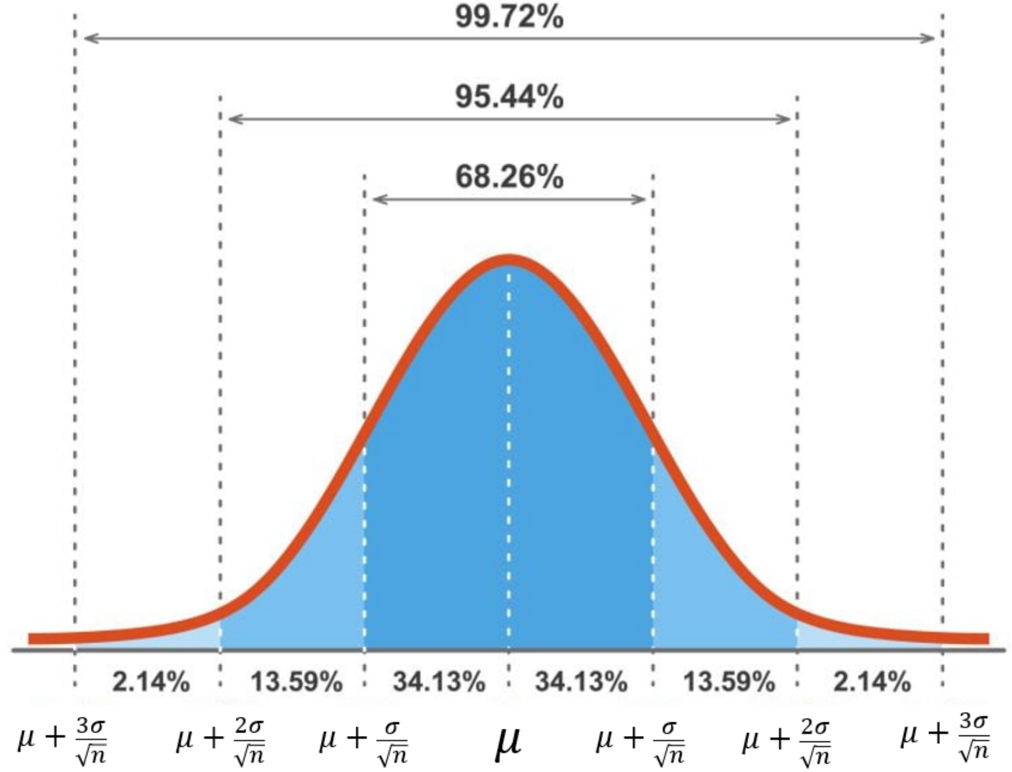
Confidence interval (CI)

An interval that has $C\%$ probability containing true mean. Where $C = 1 - \alpha$, is called confidence level, and α is called significance level.

- μ has **68%** probability within $\bar{x} \pm \frac{s}{\sqrt{n}}$
- μ has **95%** probability within $\bar{x} \pm \frac{2s}{\sqrt{n}}$
- μ has **99%** probability within $\bar{x} \pm \frac{3s}{\sqrt{n}}$

$$CI = \bar{x} \pm t_{(\frac{\alpha}{2}, n-1)} * \frac{s}{\sqrt{n}}$$

- We need to use t-distribution here because n might be smaller than 30, and only if assuming the population itself is normally distributed.



JPRF black bear

- We measured the weight from 25 black bears in JPRF forest, and got their averaged weight as 400 kg, with standard deviation 20 kg.
- Find the confidence interval for μ, with 90% confidence interval.

$$n = 25, \bar{x} = 400, s = 40, \alpha = 0.1$$

$$Upper = \bar{x} + t_{0.05,24} * \frac{s}{\sqrt{n}} = 413.7$$

$$Lower = \bar{x} - t_{0.05,24} * \frac{s}{\sqrt{n}} = 386.3$$

- Based on our sample, we have 90% confidence that the weight of a black bear in JPRF forest is between 386.3 to 413.7 kg.

What we learned

- Concept of sampling distribution
- Central limit theorem tells us
  - mean of the sampling distribution is an unbiased estimate of the population mean
  - sd of the sampling distribution is called standard error
- Standard error is not the same thing as standard deviation
- We can get a confidence interval about population mean, just by one sample!
- And keep in mind… the topic we learned in each lecture can be a course by itself

Wrap up

Before we meet again

- Review the two lectures (important)
- Review and practice with the R coding (also important)
- Maybe use what we learned to explore your own data

Next time

- Tue. 12:30, virtual on zoom
- Have a wonderful, relaxing, gull-like weekend