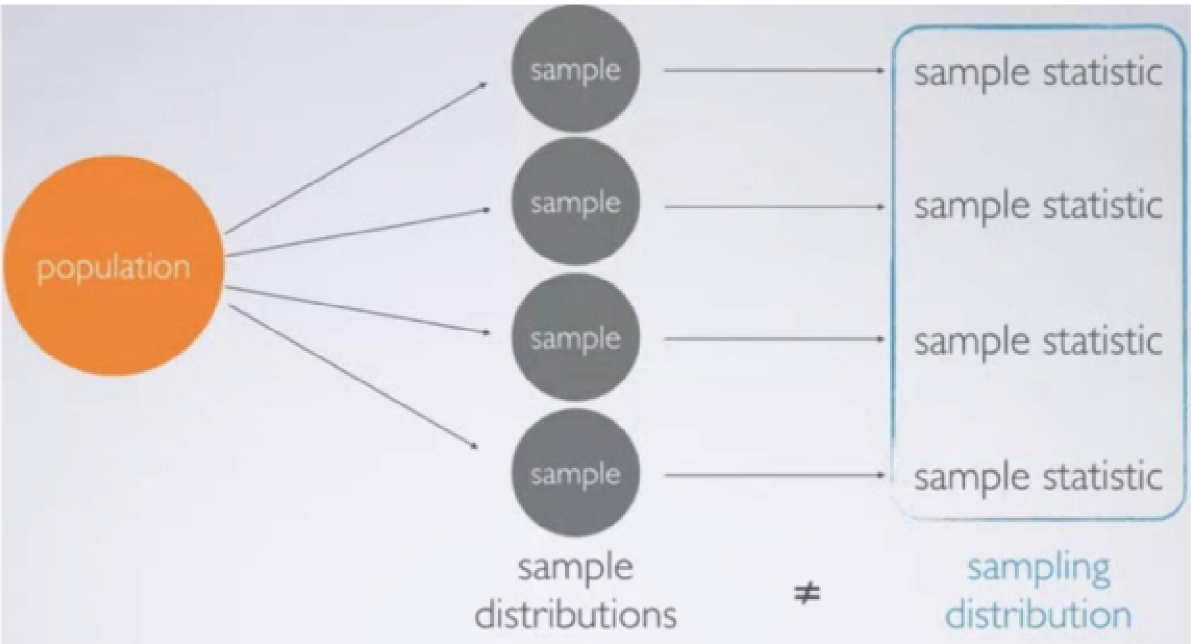
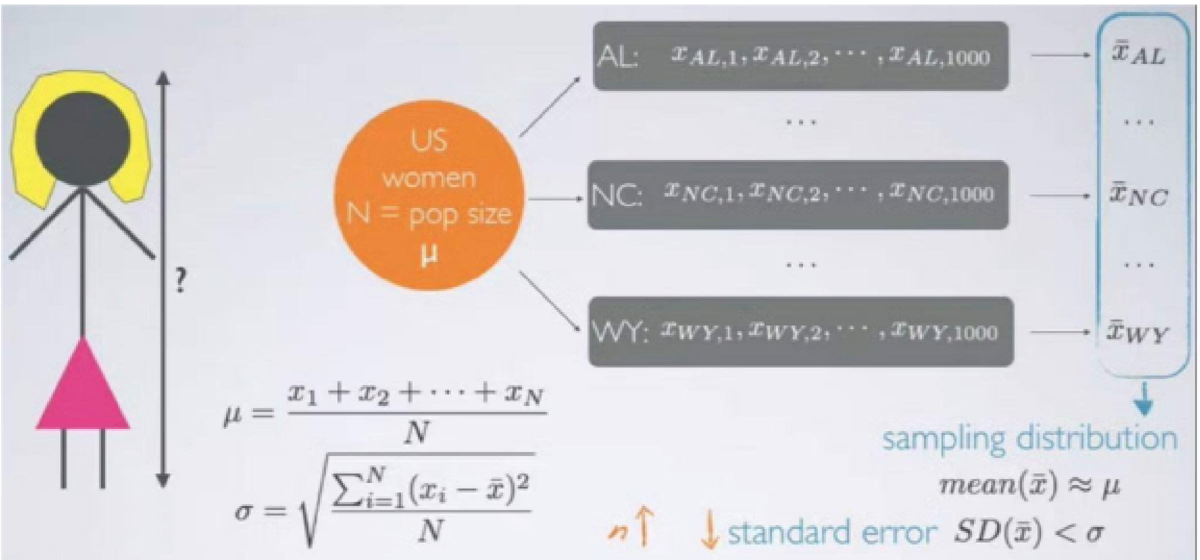
Sampling Variability



for example:



Sampling distribution: distribution of sample means, and it's roughly just like the population distribution

$\mu \approx \bar{x}$; $\sigma >> SD(\bar{x})$

if sample size $\uparrow$ $SD(\bar{x}) \downarrow$

if sample size ↓ $SD(\bar{x})$ ↑

Central Limit Theorem (CLT): The distribution of sample statistics is nearly normal, centered at the population mean, and with a standard deviation equal to the population standard deviation divided by square root of the sample size.

$$\bar{x} \sim N\left(mean = \mu, SE = \frac{\overset{s}{\cancel{\sigma}}}{\sqrt{n}}\right)$$

shape center spread

Conditions for the CLT:

1. ***Independence:*** Sampled observations must be independent.
   - random sample/assignment
   - if sampling without replacement, n < 10% of population
2. ***Sample size/skew:*** Either the population distribution is normal, or if the population distribution is skewed, the sample size is large (rule of thumb: n > 30).

if the population distribution is not normal, the more skewed the population distribution the larger sample size we need for the central limit theorem to apply.

example:

I'm about to take a trip to visit my parents and the drive is 6 hours. I make a random playlist of 100 songs. What is the probability that my playlist lasts the entire drive?

6 hours = 360 minutes

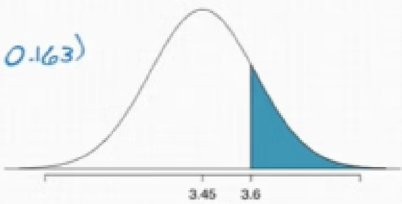
$P(X_1 + X_2 + \dots + X_{100} > 360 \text{ min}) = ?$

$P(\bar{X} > 3.6) = ?$

$\bar{X} \sim N\left(mean = \mu = 3.45, SE = \frac{\sigma}{\sqrt{n}} = \frac{1.63}{\sqrt{100}} = 0.163\right)$

$Z = \frac{3.6 - 3.45}{0.163} = 0.92$

$P(Z > 0.92) = 0.179$



Four plots: Determine which plot (A, B, or C) is which.
(1) The distribution for a population ($\mu = 10, \sigma = 7$),
(2) a single random sample of 100 observations from this population,
(3) a distribution of 100 sample means from random samples with size 7, and
(4) a distribution of 100 sample means from random samples with size 49.

Confidence Interval

A plausible range of values for the population parameter.

$$\bar{x} \pm \underbrace{z^* \cdot \frac{s}{\sqrt{n}}}_{\text{Margin}} \quad \text{SE}$$

accuracy is defined in terms of whether or not the confidence interval contains the true population parameter.

and precision refers to the width of a confidence interval.

note: confidence interval is not about individuals in the population but instead about the true population parameter.

$$ME = z \frac{s}{\sqrt{n}} \Rightarrow n = \left(\frac{z\,s}{ME}\right)^2$$

example:

$ME \leq 4$ pts

$CI = 90\% \Rightarrow z = 1.65$ $\qquad n = 55.13 \sim 56$

$\sigma = 18$ pts

now:

$$\frac{ME}{2} = z \frac{s}{\sqrt{4n}}$$

example 2:

The General Social Survey asks: "For how many days during the past 30 days was your mental health, which includes stress, depression, and problems with emotions, not good?" Based on responses from 1,151 US residents, the survey reported a 95% confidence interval of 3.40 to 4.24 days in 2010. Interpret this interval in context of the data.

We are 95% confident that Americans on average have 3.40 to 4.24 bad mental health days per month.

In this context, what does a 95% confidence level mean?

95% of random samples of 1,151 Americans will yield CIs that capture the true population mean of number of bad mental health days per month.

example 3:

A sample of 50 college students were asked how many exclusive relationships they've been in so far. The students in the sample had an average of 3.2 exclusive relationships, with a standard deviation of 1.74. In addition, the sample distribution was only slightly skewed to the right. Estimate the true average number of exclusive relationships based on this sample using a 95% confidence interval.

1. random sample & 50 < 10% of all college students
We can assume that the number of exclusive relationships one student in the sample has been in is independent of another.
2. $n > 30$ & not so skewed sample
We can assume that the sampling distribution of average number of exclusive relationships from samples of size 50 will be nearly normal.

$n = 50$
$\bar{x} = 3.2$
$s = 1.74$

$$SE = \frac{s}{\sqrt{n}} = \frac{1.74}{\sqrt{50}} = 0.246$$

$$\bar{x} \pm z \cdot SE = 3.2 \pm 0.48 = (2.72, 3.68)$$

We are 95% confident that college students on average have been in 2.72 to 3.68 exclusive relationships.

Last modified: 7:00 PM