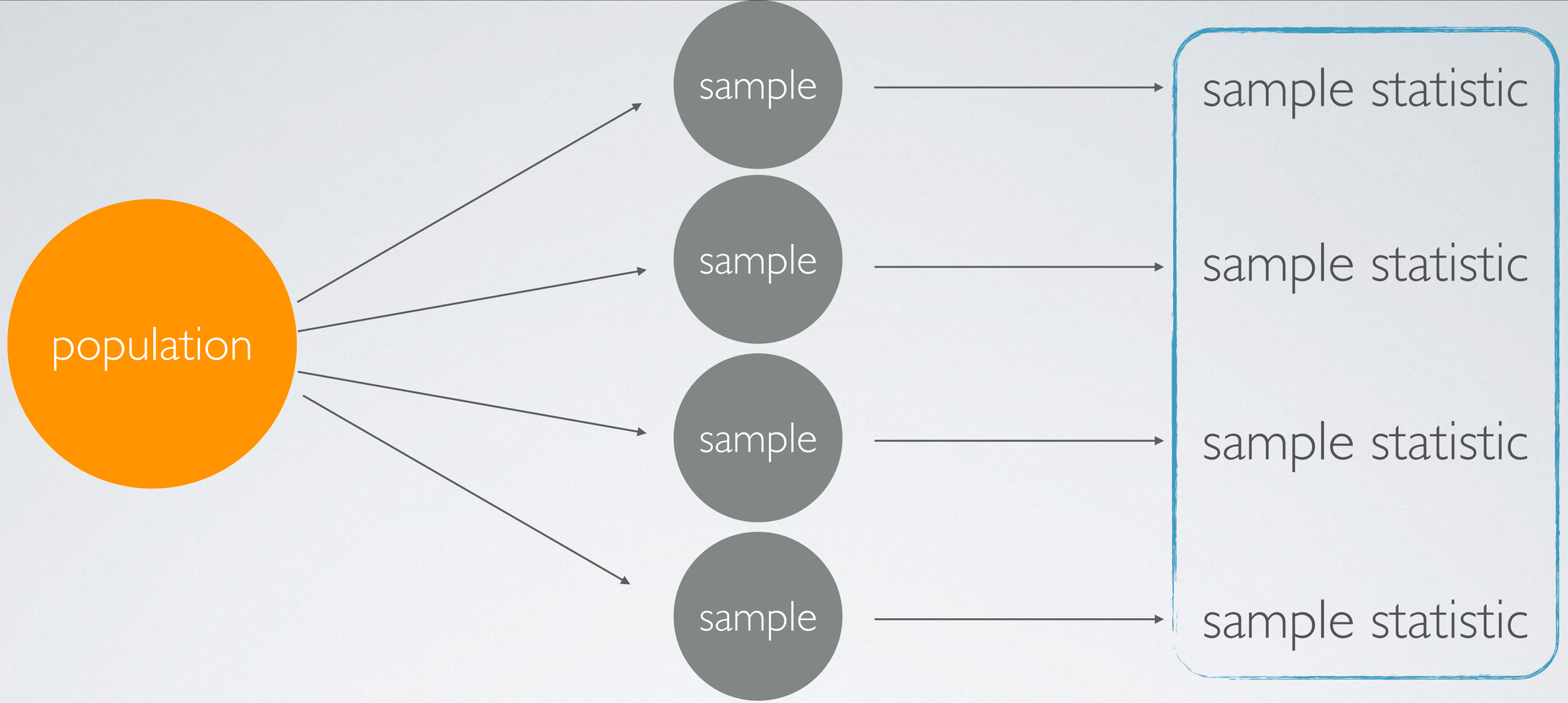
sampling variability & CLT

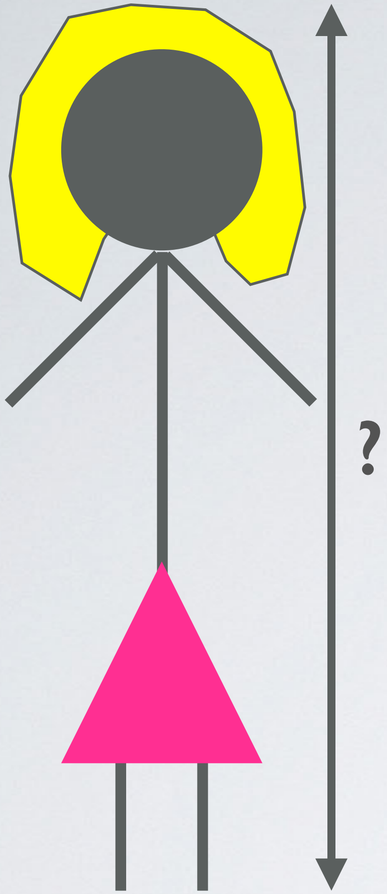
- sampling distribution
- CLT + conditions for CLT
- demo

Duke UNIVERSITY

Dr. Mine Çetinkaya-Rundel
Duke University



sample distributions ≠ sampling distribution



US
women
N = pop size
μ

AL: $x_{AL,1}, x_{AL,2}, \cdots, x_{AL,1000}$

NC: $x_{NC,1}, x_{NC,2}, \cdots, x_{NC,1000}$

WY: $x_{WY,1}, x_{WY,2}, \cdots, x_{WY,1000}$

$\bar{x}_{AL}$ ... $\bar{x}_{NC}$ ... $\bar{x}_{WY}$

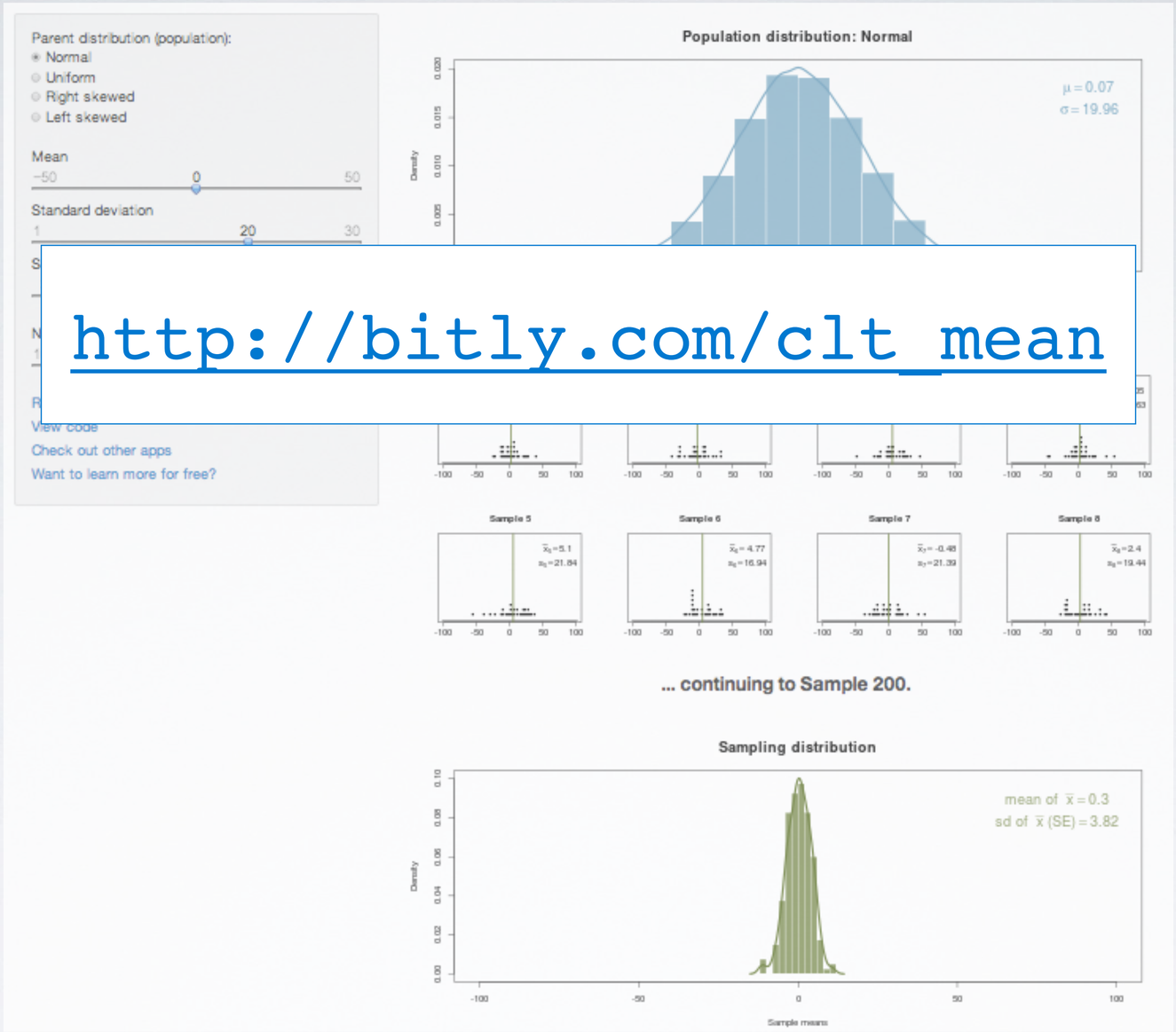
sampling distribution

$$\mu = \frac{x_1 + x_2 + \cdots + x_N}{N}$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \bar{x})^2}{N}}$$

$mean(\bar{x}) \approx \mu$

$n \uparrow$ $\downarrow$ standard error $SD(\bar{x}) < \sigma$



http://bitly.com/clt_mean

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.

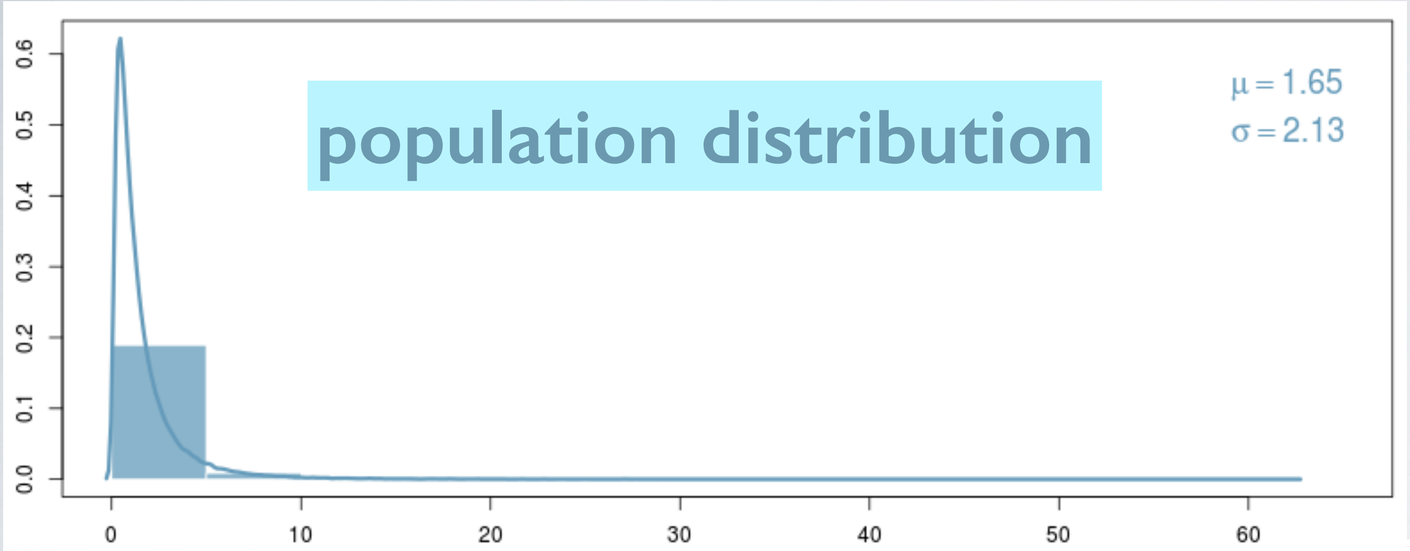
$$\bar{x} \sim N\left(mean = \mu, SE = \frac{\sigma}{\sqrt{n}}\right)$$

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).


focus on:
10% condition —
"if sampling without
replacement,
n <10% of the
population"



focus on:
sample size /
skew condition