

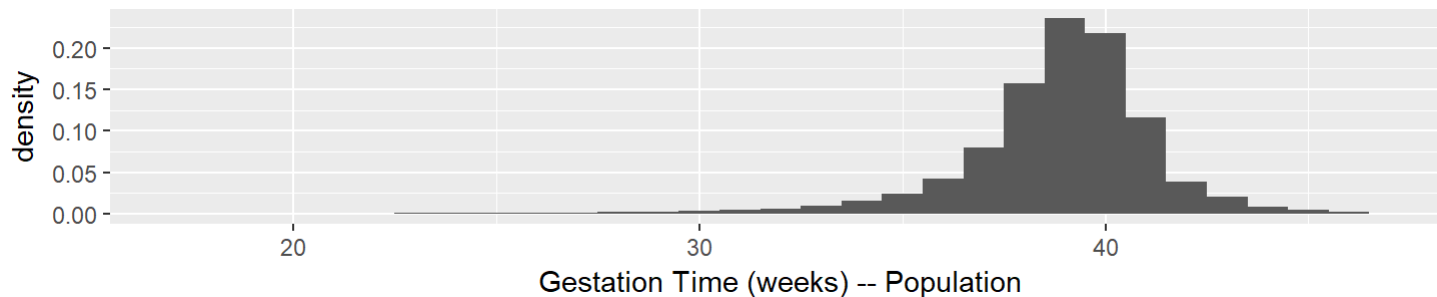
# Population Distribution, Sample Distribution, Sampling Distribution, and Confidence Intervals

Evan L. Ray

November 17, 2017

# Distribution of the Population

- For each possible gestation time, what proportion of babies in the population had that gestation time?

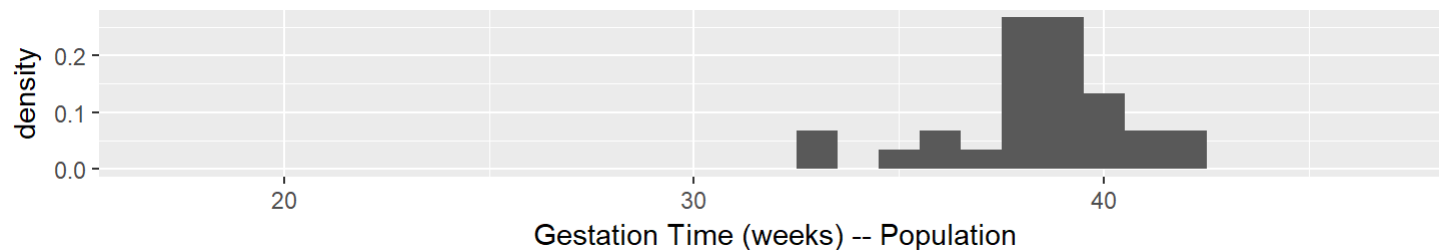


- Population mean: 38.8 weeks
- Population standard deviation: 2.6 weeks
- About 95% of babies in the population had gestation times between  $(38.8 - 2 * 2.6)$  weeks and  $(38.8 + 2 * 2.6)$  weeks

# Distribution of a Sample

- For each possible length of gestation time, what proportion of babies in the **sample** had that gestation time?

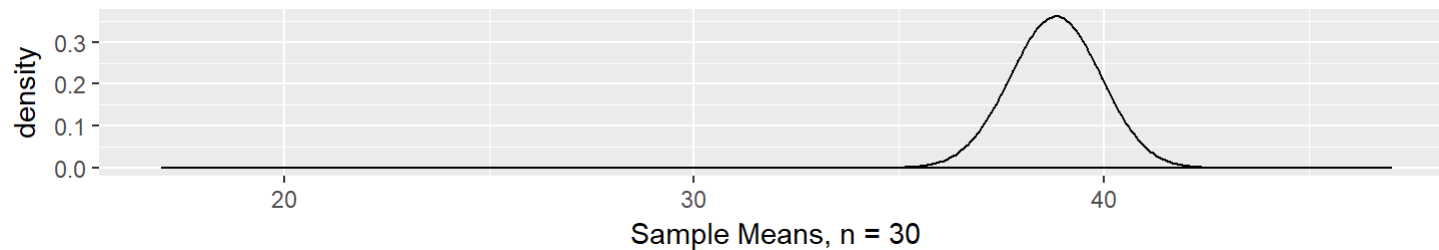
```
babies_sample <- sample_n(babies, size = 30)
```



- Sample mean: 38.7 weeks
- Sample standard deviation: 2.2 weeks
- About 95% of babies in the sample had gestation times between  $(38.7 - 2 * 2.2)$  weeks and  $(38.7 + 2 * 2.2)$  weeks

# Sampling Distribution of Sample Mean

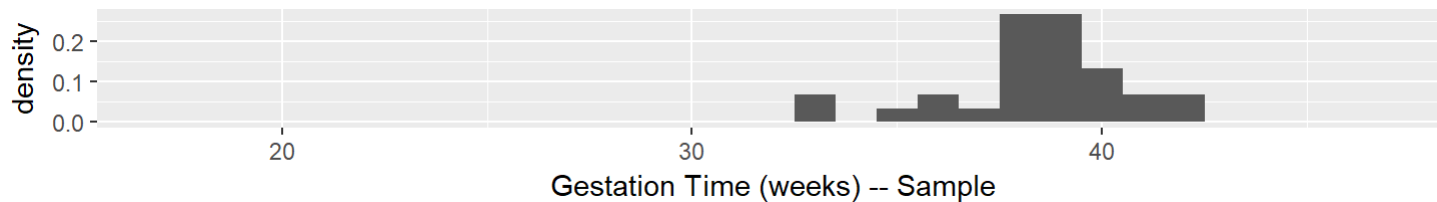
- The **sampling distribution** is the distribution of values of the sample mean, across all different samples of a certain size  $n$ .
- If  $n$  is large enough,  $\bar{X} \sim \text{Normal}(\mu, \sigma/\sqrt{n})$



- Population mean: 38.8 weeks
- Population standard deviation: 2.6 weeks
- About 95% of samples of size 30 have sample mean gestation times between  $(38.8 - 2 * \frac{2.6}{\sqrt{30}})$  and  $(38.8 + 2 * \frac{2.6}{\sqrt{30}})$

# 95% Conf. Interval for Population Mean

- (best guess of population mean)  $\pm$  (margin of error)
- $\bar{x} \pm 2s/\sqrt{n}$



- Sample mean: 38.7 weeks
- Sample standard deviation: 2.2 weeks
- We are "95% Confident" that the population mean gestation time is between  $(38.7 - 2 * \frac{2.2}{\sqrt{30}})$  and  $(38.7 + 2 * \frac{2.2}{\sqrt{30}})$
- "95% Confident" means: 95% of intervals constructed this way from different samples will contain the population mean