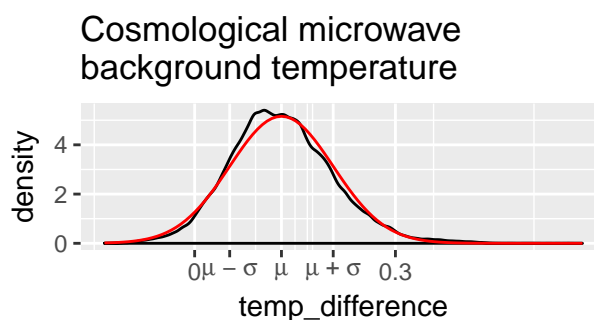
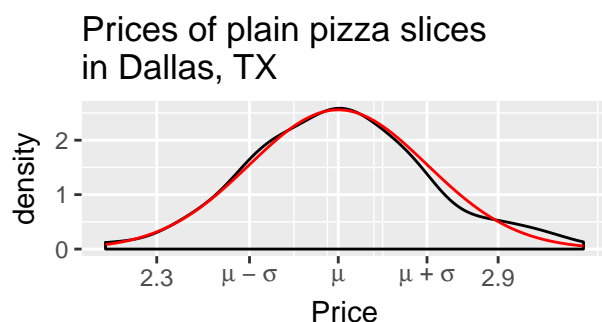
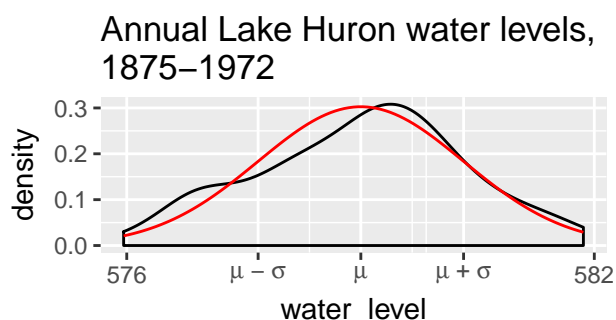
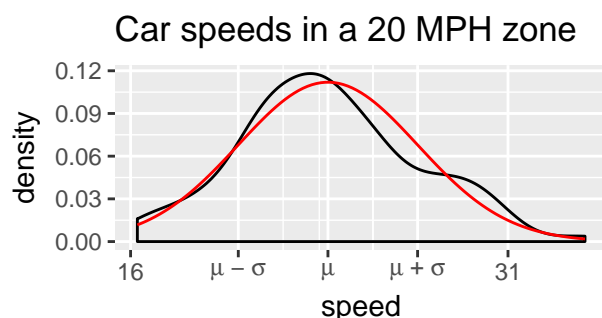


The Normal Distribution and the Sampling Distribution of the Sample Mean

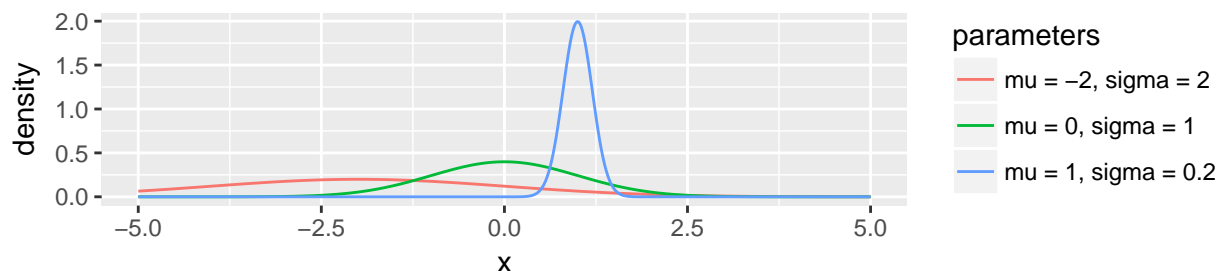
The Normal Model (Bell Curve) - See Chapter 5

In each of the plots below,

- the black curve is a density plot of the observed sample data
- the red curve is a density curve for a normal model with the same mean and standard deviation as the sample data



- Let the random variable X denote the numeric value of one of these variables for a randomly sampled item from the population
 - Example: X is the speed of a randomly selected car driving in a 20 MPH speed zone.
 - Example: X is the price of a randomly selected slice of pizza in Dallas, TX.
- We can **model** the value of X as being a draw from a normal distribution
- $X \sim \text{Normal}(\mu, \sigma)$
- Read: “ X is modeled as following a normal distribution with mean μ and standard deviation σ ”
- μ and σ are **parameters** that control the center and spread of the distribution

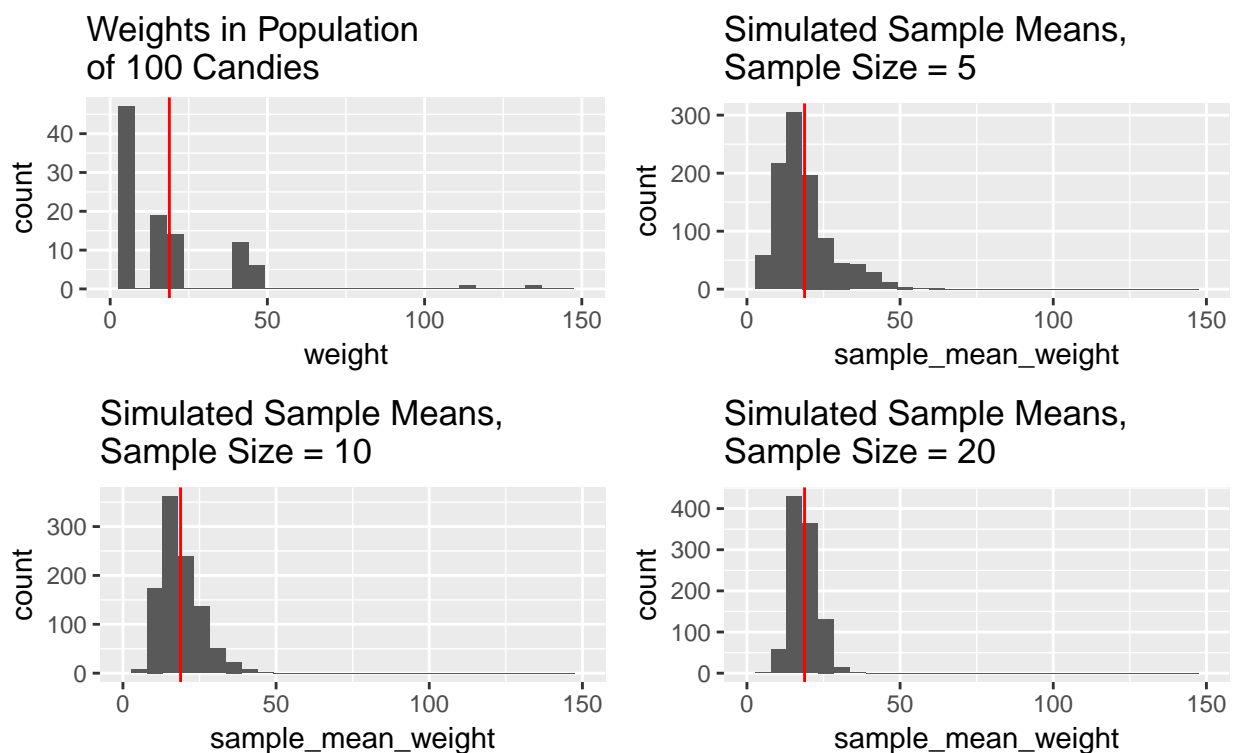


The Sampling Distribution of the Sample Mean - See Chapter 17

Our Goal: Use the sample mean to estimate the population mean. **What can we say about how close the sample mean should be to the population mean?**

Sampling Distribution of the Sample Mean: The distribution of values of the sample mean that we would obtain from all possible samples of a certain size n .

Last class: we took many different samples from a population of candy bars, and calculated the sample mean for each of those different samples:



Key Observations About the Sampling Distributions:

- **Center:** The average of the sample means across different samples is close to the population mean
- **Spread:** As the sample size increases, there is less spread in the sample means from different samples
- **Shape:** As the sample size increases, the distribution of sample means from different samples becomes more symmetric, less skewed.

Central Limit Theorem

- Y_1, Y_2, \dots, Y_n are **independent** observations of a **quantitative variable**
- Population has mean μ and standard deviation σ
- Compute the sample mean: $\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i$
- The sampling distribution of \bar{Y} :
 - has mean μ
 - has standard deviation σ/\sqrt{n}
 - for **large enough** n , it is approximately normal
 - putting this together: If n is large enough, $\bar{Y} \sim \text{Normal}(\mu, \sigma/\sqrt{n})$ (approximately).
- “How large does n have to be?” Depends on shape of population distribution - but if the distribution is symmetric enough that you are calculating a mean, about 30 is usually enough.

Example (Adapted from SDM4 17.62)

Farmers measure daily milk production in pounds. Ayrshire cows average 55 pounds of milk per day, with a standard deviation of 6 pounds. Suppose I plan to select 100 Ayrshire cows at random and measure their milk production, and then compute the average milk production for those 100 cows (I know that the mean will be a reasonable summary of the center of this distribution). What is the sampling distribution for the sample mean?

(a) **Examine conditions/assumptions for using the central limit theorem to say something about the sampling distribution of the sample mean.**

- Is the milk production for different cows **independent**?
- Am I thinking about a **quantitative variable**?
- Is **calculating a mean reasonable**/is my **sample size** large enough?

(b) **What is the sampling distribution of the sample mean, approximately?**