15 - Sampling distributions

On a scale of 1-10, how much do you like pineapple on pizza?

Our news channel surveyed **one** people. Here are the results:

FOX	ABC	NBC	PBS
1	5	6	2

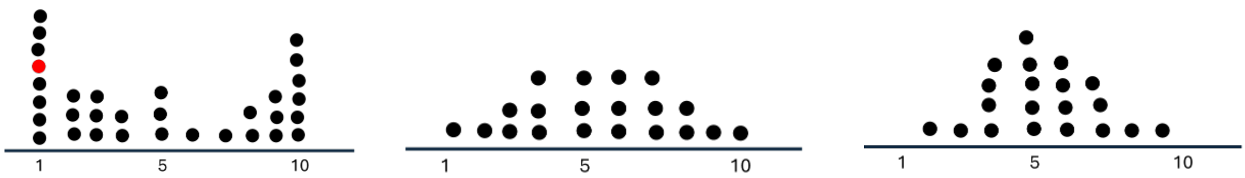
Our news channel surveyed **three** people. Here's their average:

FOX	ABC	NBC	PBS
3	5.2	6.1	4

Our news channel surveyed **five** people. Here's their average:

FOX	ABC	NBC	PBS
4.5	5.9	5.2	4.8

Moral: Average of a large SRS is often closer to the true population mean.



Moral 2: The distribution of average of samples of size n approaches normal distribution as $n \to \infty$.

This distribution is called the **sampling distribution** of $\bar{x}$ of size n.

10% Condition. We sample without replacement via SRS. If the population size N is at least 10 times the sample size n, then sampling without replacement is approximately like sampling with replacement.

Facts about sampling distribution of $\bar{x}$:
Suppose $\bar{x}$ is the mean of an SRS of size n drawn from a population with mean μ and standard deviation σ. Then this sampling distribution of $\bar{x}$ has:
- Mean μ
- Standard deviation $\frac{\sigma}{\sqrt{n}}$ if the 10% condition $N \geq 10n$ holds.

- Normal if the population distribution is normal.
- Approximately Normal if the population distribution is not too skewed and n≥30. (Central Limit Theorem)

Example 1. Studies find the minimum sulfur compound DMS odor threshold of adults follows a distribution with pop. mean 25 mg/L and pop. sd 7 mg/L. We take an SRS of 10 adults and determine the mean odor threshold $\bar{x}$ for the individuals in the sample.

(a) What is the mean of the sampling distribution of $\bar{x}$? 25 mg/L

(b) What is the standard deviation of the sampling distribution of $\bar{x}$?

Population has $\geq 10n = 10 \cdot 10 = 100$ adults so we can use sd formula: $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{7}{10} \approx 2.2$

Example 2. The height of young women follows a Normal distribution with mean μ = 64.5 inches and standard deviation σ = 2.5 inches.

(a) Probability a randomly selected young woman is taller than 66.5 inches:

```
pnorm(66.5, mean=64.5, sd = 2.5, lower.tail=FALSE) =
pnorm((66.5-64.5)/2.5, lower.tail=FALSE) = 0.788 = 78.8%
```

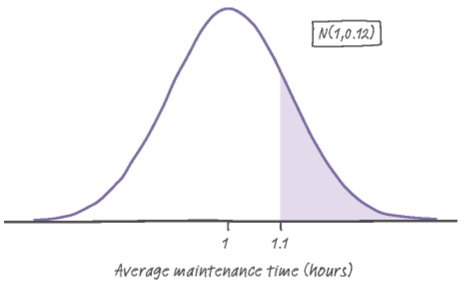
(b) Probability the mean height of an SRS of 10 young women exceeds 66.5 inches:

Pop. has $N \geq 10n = 100$ young women so $\sigma_{\bar{x}} = \frac{2.5}{\sqrt{10}} \approx 0.79$

$$P(\bar{x} > 66.5) = \text{pnorm}\left(\frac{66.5 - 64.5}{0.79}, \text{lower.tail=FALSE}\right) = 0.0057.$$

So it is very unlikely (< 1% chance) that an SRS of 10 young women has average height exceeding 66.5 inches.

Example 3. Your company has a contract to perform maintenance on thousands of air-conditioning units. Based on service records from the past year, the time (in hours) that a technician requires to complete the work follows a strongly right-skewed distribution with μ = 1 hour and σ= 1 hour. This week, your company will service an SRS of 70 AC units in the city. You plan to budget an average of 1.1 hours per unit for a technician to complete the work. Will this be enough? Find the probability the average maintenance time for 70 units exceeds 1.1 hours.



Facts about sampling distribution of proportion $\hat{p}$:

Choose an SRS of size n from a population of size N with proportion p of successes. Let $\hat{p}$ be the sample proportion of successes. Then this sampling distribution of has:

- Mean $\hat{p}$
- Standard deviation $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$ if the 10% condition $N \geq 10n$ holds.
- Approximately Normal if the Large Counts Condition is satisfied:

$$np \geq 10 \text{ and } n(1-p) \geq 10$$

Example 4. A poll finds: of an SRS of 1500 first-year college students, 35% of all first-year students attend college within 50 miles of home. Find the probability that the random sample of 1500 students will give a result within 2% of the true value.