Activity.

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

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

FOX ABC NBC PBS

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

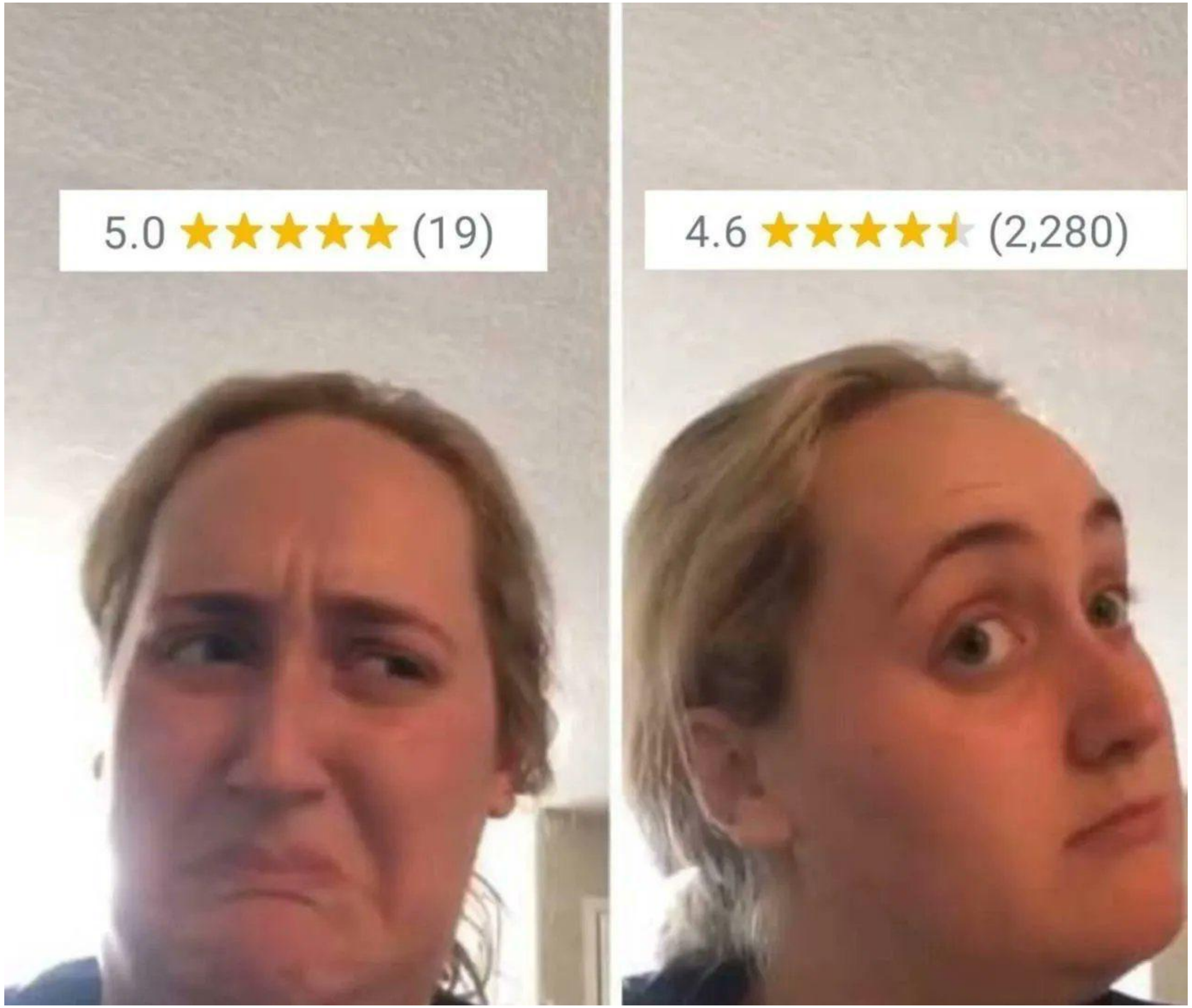
FOX ABC NBC PBS

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

FOX ABC NBC PBS

Pattern 1.

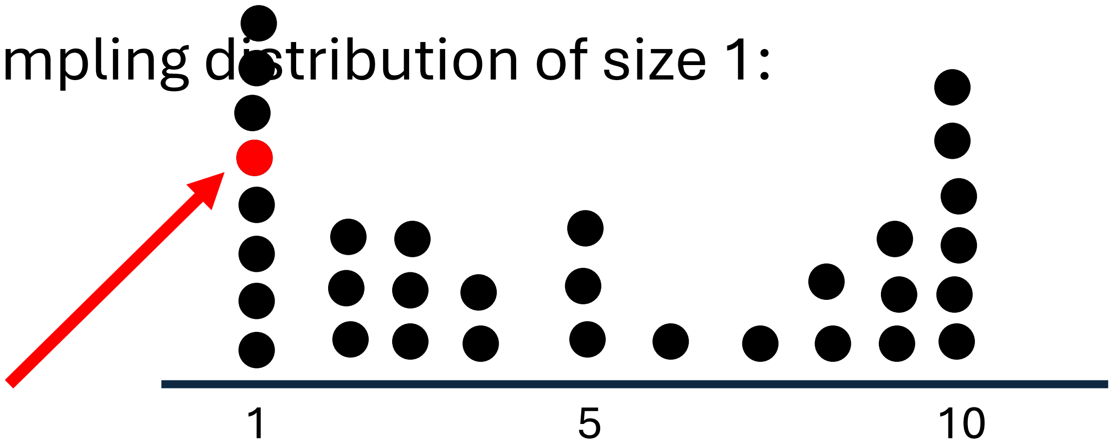
Average of a large SRS is often less extreme and closer to the true population mean.


5.0 ★★★★★ (19)
4.6 ★★★★½ (2,280)

Our news channel surveyed **one** people. Sampling distribution of size 1:

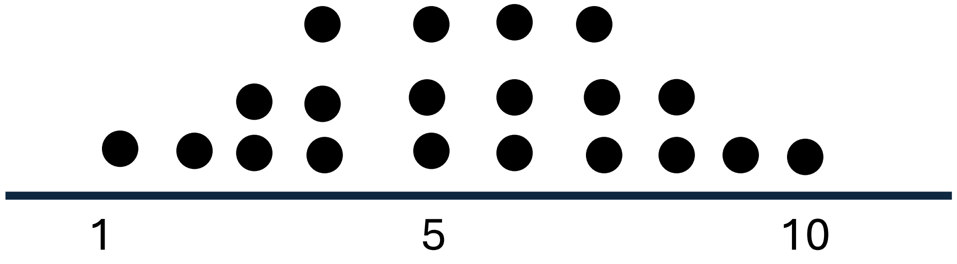
1	5	6	2	10
FOX	ABC	NBC	PBS	CBC

One dot = one person's opinion



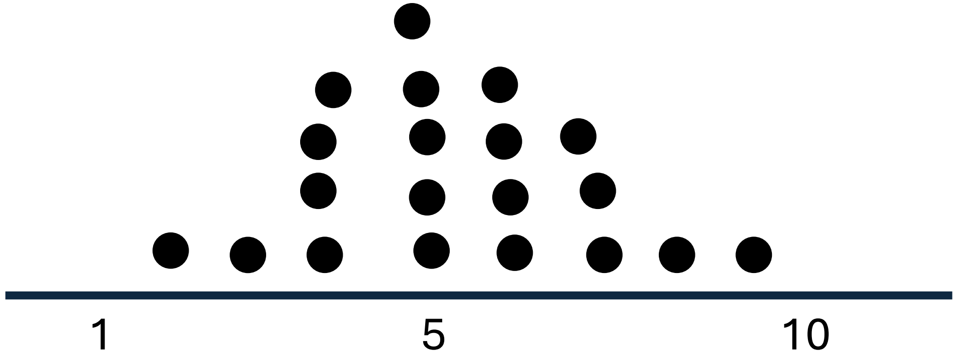
Our news channel surveyed **three** people. Sampling distribution of size 3:

3	5.2	6.1	4	7
FOX	ABC	NBC	PBS	CBC



Our news channel surveyed **ten** people. Sampling distribution of size 10:

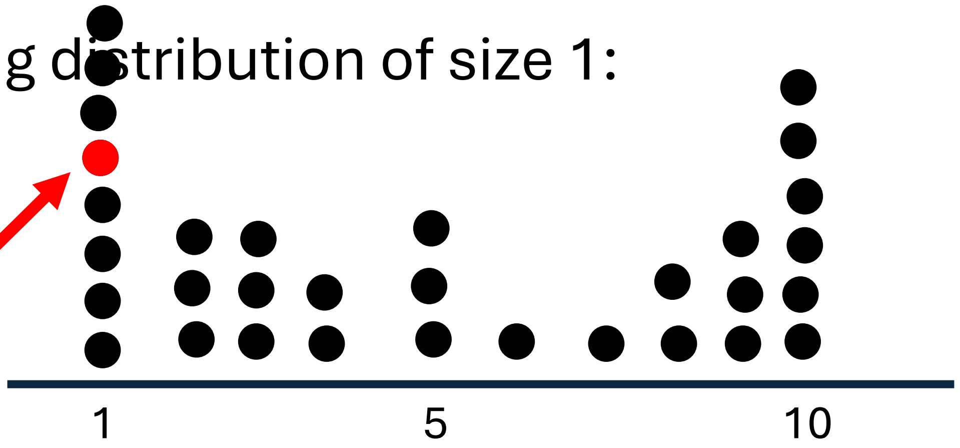
4.5	5.9	5.2	4.8	6.1
FOX	ABC	NBC	PBS	CBC



Our news channel surveyed **one** people. Sampling distribution of size 1:

1 FOX	5 ABC	6 NBC	2 PBS	10 CBC

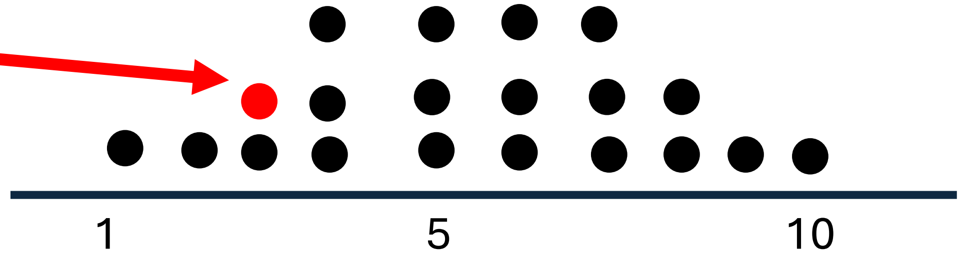
One dot = one person's opinion



Our news channel surveyed **three** people. Sampling distribution of size 3:

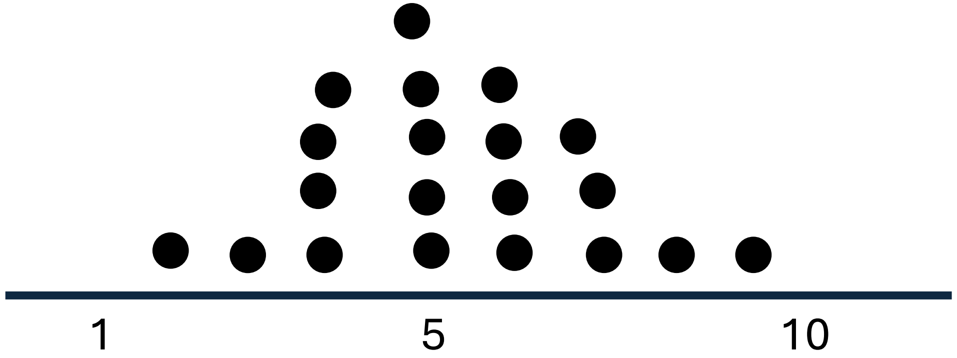
One dot = 3 people's opinions, averaged

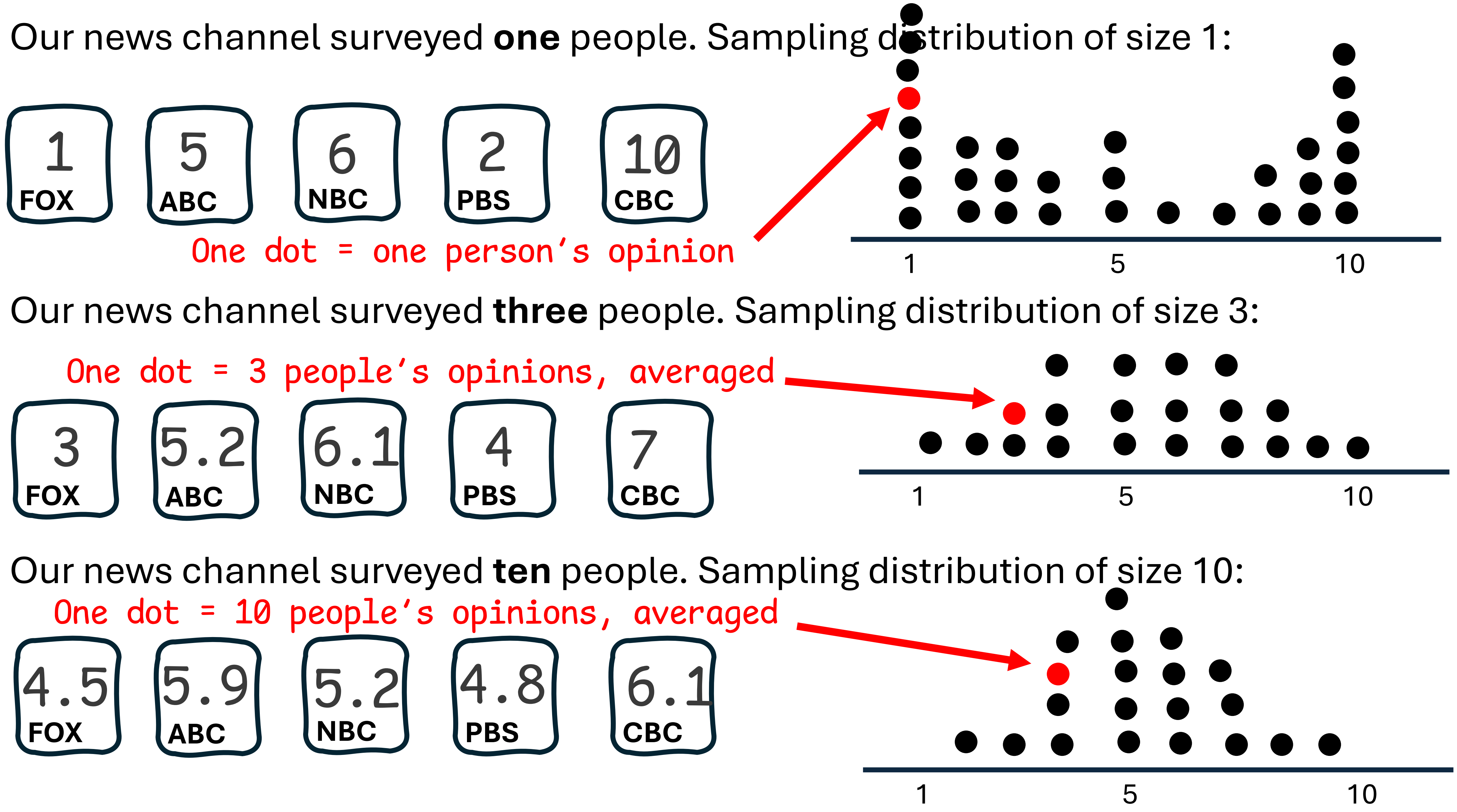
3 FOX	5.2 ABC	6.1 NBC	4 PBS	7 CBC



Our news channel surveyed **ten** people. Sampling distribution of size 10:

4.5 FOX	5.9 ABC	5.2 NBC	4.8 PBS	6.1 CBC




Our news channel surveyed **one** people. Sampling distribution of size 1:
1 FOX
5 ABC
6 NBC
2 PBS
10 CBC
One dot = one person's opinion
1
5
10
Our news channel surveyed **three** people. Sampling distribution of size 3:
One dot = 3 people's opinions, averaged
3 FOX
5.2 ABC
6.1 NBC
4 PBS
7 CBC
1
5
10
Our news channel surveyed **ten** people. Sampling distribution of size 10:
One dot = 10 people's opinions, averaged
4.5 FOX
5.9 ABC
5.2 NBC
4.8 PBS
6.1 CBC
1
5
10

https://onlinestatbook.com/stat_sim/sampling_dist/

Pattern 2: **Central Limit Theorem (CLT).**

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

Pattern 2: **Central Limit Theorem (CLT).**

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

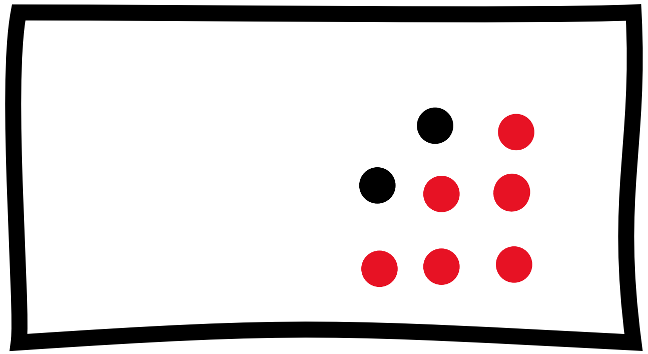
This distribution is called the **sampling distribution** of $\bar{x}$ of size n.

Sampling with replacement vs Sampling without replacement

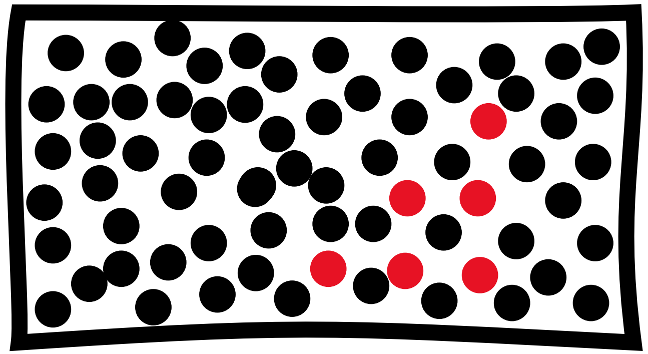
10% Condition.
Suppose we sample without replacement via SRS (simple random sample).
If the population size N is at least 10 times the sample size n, then sampling without replacement is approximately like sampling with replacement.

Sampling with replacement vs Sampling without replacement

10% Condition.
Suppose we sample without replacement via SRS (simple random sample). If the population size N is at least 10 times the sample size n, then sampling without replacement is approximately like sampling with replacement.



When $n \approx N$, it's likely to interview the same person twice, so sampling without replacement $\neq$ sampling with replacement.



When $n \leq \frac{1}{10}N$, it's unlikely to interview the same person twice, so sampling without replacement $\approx$ 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.
- Normal if the the population distribution is not too skewed and $n \geq 30$.
  (Central Limit Theorem)

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

Why? If X_i is RV of ith person, then the sample mean is a RV $X = \frac{1}{n}(X_1 + X_2 + \cdots + X_n)$:

$$\mathrm{E}(X) = \frac{1}{n} E(X_1 + \cdots + X_n) = \frac{1}{n} \cdot (n\mu),$$

$$\sigma_X^2 = \frac{1}{n^2} \sigma_{X_1 + \cdots + X_n}^2 = \frac{1}{n^2}(\sigma_{X_1}^2 + \cdots + \sigma_{X_n}^2) = \frac{1}{n^2}(n\sigma^2) \quad \Rightarrow \quad \sigma_X = \frac{\sigma}{\sqrt{n}}$$

X_i are roughly independent RVs because the 10% Condition holds.

Sample Mean Example 1

Sulfur compounds such as dimethyl sulfide (DMS) are sometimes present in wine. DMS causes "off-odors" in wine, so winemakers want to know the odor threshold, the lowest concentration of DMS that the human nose can detect. Extensive studies have found that the DMS odor threshold of adults follows a distribution with mean $\mu = 25$ micrograms per liter and standard deviation $\sigma = 7$ micrograms per liter. Suppose we take an SRS of 10 adults and determine the mean odor threshold $\bar{x}$ for the individuals in the sample.

PROBLEM:

(a) What is the mean of the sampling distribution of $\bar{x}$? Explain.

(b) What is the standard deviation of the sampling distribution of $\bar{x}$? Check that the 10% condition is met.

Sample Mean Example 2

PROBLEM: The height of young women follows a Normal distribution with mean $\mu = 64.5$ inches and standard deviation $\sigma = 2.5$ inches.

(a) Find the probability that a randomly selected young woman is taller than 66.5 inches. Show your work.

(b) Find the probability that the mean height of an SRS of 10 young women exceeds 66.5 inches.

Sample Mean Example 3

Your company has a contract to perform preventive maintenance on thousands of air-conditioning units in a large city. 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 $\mu = 1$ hour and $\sigma = 1$ hour. In the coming week, your company will service an SRS of 70 air-conditioning 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?

PROBLEM: What is the probability that the average maintenance time $\bar{x}$ for 70 units exceeds 1.1 hours? Show your work.

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 $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$
- Approximately Normal if the Large Counts Condition is satisfied:

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

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 $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$
- 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.

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 $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$
- 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.