6.1 Sampling Distributions

Dr. Lauren Perry

Goals

1. Find the distribution of a sample mean.
2. Estimate probabilities for a sample mean.

Sampling Error

We want to use a sample to learn something about a population, but no sample is perfect!

Sampling error is the error resulting from using a sample to estimate a population characteristic.

Sampling Error

If we use a sample mean $\bar{x}$ to estimate μ, chances are that $\bar{x} \neq \mu$ (they might be close but... they might not be!). We will consider

- How close *is* $\bar{x}$ to μ?
- What if we took many samples and calculated $\bar{x}$ many times?
  - How would that relate to μ?
  - What would be the distribution of these values?

Sampling Distribution

The distribution of a statistic (across all possible samples of size n) is called the **sampling distribution**.

For a variable x and given a sample size n, the distribution of $\bar{x}$ is called the **sampling distribution of the sample mean** or the **distribution of $\bar{x}$**.

Example

Suppose our population is the five starting players on a particular basketball team. We are interested in their heights (measures in inches). The full population data is

Player	A	B	C	D	E
Height	76	78	79	81	86

The population mean is $\mu = 80$.

Example

Consider all possible samples of size $n = 2$:

Sample	A,B	A,C	A,D	A,E	B,C	B,D	B,E	C,D	C,E	D,E
$\bar{x}$	77	77.5	78.5	81.0	78.5	79.5	82.0	80.0	82.5	83.5

There are 10 possible samples of size 2.

- Of these samples, 10% have means exactly equal to μ.
  - For a *random* sample of size 2, you'd have a 10% chance to find $\bar{x} = \mu$.

In general, the larger the sample size, the smaller the sampling error tends to be in estimating μ using $\bar{x}$.

In practice, we have one sample and μ is unknown.

The sampling distribution of the sample mean

For the distribution of $\bar{X}$

- The mean of the distribution is $\mu_{\bar{X}} = \mu$.
- The standard deviation is $\sigma_{\bar{X}} = \sigma/\sqrt{n}$.

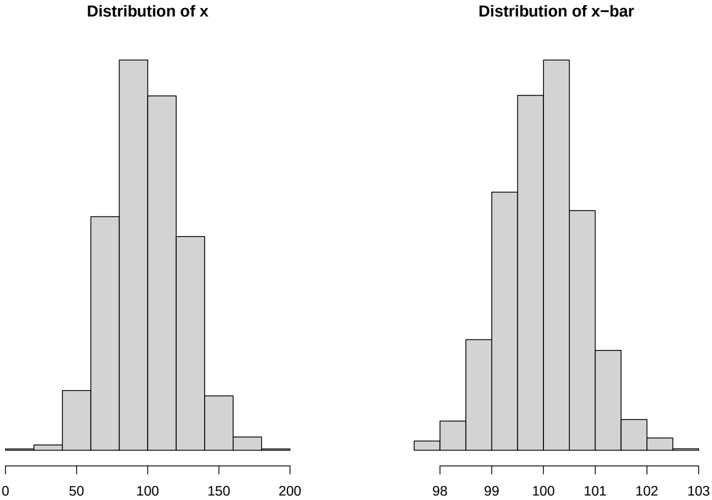
We refer to the standard deviation of a sampling distribution as **standard error**.

Checkpoint

The mean living space for a detached single family home in the United States is 1742 ft^2 with a standard deviation of 568 square feet. For samples of 25 homes, determine the mean and standard error of $\bar{x}$.

The Distribution of $\bar{X}$

The plots show (A) a random sample of 1000 from a Normal(100, 25) distribution and (B) the approximate sampling distribution of $\bar{X}$ when X is Normal(100, 25).



The Distribution of $\bar{X}$

In fact, if X is Normal(μ, σ), then $\bar{X}$ is Normal($\mu_{\bar{X}} = \mu$, $\sigma_{\bar{X}} = \sigma/\sqrt{n}$).

Surprisingly, we see a similar result for $\bar{X}$ even when X is not normally distributed!

The Central Limit Theorem

For relatively large sample sizes, the random variable $\bar{X}$ is approximately normally distributed *regardless of the distribution of* X:

$$\bar{X} \text{ is Normal}(\mu_{\bar{X}} = \mu, \sigma_{\bar{X}} = \sigma/\sqrt{n}).$$

Notes

- This approximation improves with increasing sample size.
- In general, "relatively large" means sample sizes $n \geq 30$.