Chapter 8

- Sample proportion
- Sample mean
- True proportion (population)
- Sample proportion (sample)

Unbiased

Center of the sampling distribution is equal to the true population parameter.

Variability — spread of its sampling distribution
Goal: low bias. low variability.

Reduce variability → increase sample size

What do we expect of sample proportion?

1. The values of $\hat{p}$ vary from sample to sample in a predictable way.

2. When the sample size $n$ is large, the sample proportion can take on many possible values in the range of 0 to 1, so the random variable $\hat{p}$ can be viewed as a continuous random variable with a density curve as its model.

3. When the sample size is large, the distribution of $\hat{p}$ can be modeled approximately with a normal distribution.
4. The center of the distribution of \( \hat{p} \) is at the true proportion \( p \) (for any sample size \( n \) and any value of \( p \)).

5. With a larger sample size \( n \), \( \hat{p} \) tends to be closer to the true proportion \( p \). Thus, \( \hat{p} \) vary less around \( p \). The variability also depends on the value of \( p \). The variation of \( \hat{p} \) depends on \( \sqrt{\frac{n}{p(1-p)}} \).

Standard deviation of \( \hat{p} \):

\[
\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}
\]

- \( n = 50 \), \( p = 0.5 \)

\[
\sigma_{\hat{p}} = \sqrt{\frac{0.5(1-0.5)}{50}} = \sqrt{\frac{0.5 \times 0.5}{50}} = 0.07
\]
Sample distribution of $\hat{p}$

1. $E\hat{p} = \hat{p}$
   
   $\hat{p}$ is an **unbiased** estimator of $p$.

2. $\sqrt{\frac{p(1-p)}{n}} \quad n \uparrow \Rightarrow \sigma_{\hat{p}} \downarrow$

3. If $n$ is sufficiently large, the distribution of $\hat{p}$ eventually looks like a normal distribution with mean $\mu_\hat{p}$ and standard derivation $\sigma_{\hat{p}}$.

$\hat{p}$ approx. $N(\mu_\hat{p}, \sigma_{\hat{p}}) = N(p, \sqrt{\frac{p(1-p)}{n}})$
The standard error of \( \hat{p} \) is
\[
\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}.
\]

If \( np \geq 5 \) and \( n(1-p) \geq 5 \), then we say the sample is sufficiently large.

**Example:** A research believes that 2% of females wear Eternity perfume. Suppose that the research is correct. What is the probability that the proportion of Eternity wearers in a sample of 510 females would be less than 1%?

\[ p = 0.02 \]
\[ n = 510 \]
\[ P(\hat{p} < 0.01) \]
\[ np = (0.02 \times 510) = 10.2 > 5 \]
\[ n(1-p) = (0.98 \times 510) = 501.8 > 5 \]
\[ \hat{p} = \hat{p} = 0.02 \]
\[ \sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.2 \times 0.88}{510}} = 0.0062 \]
\[ \hat{p} \sim N(0.02, 0.0062) \]
\[ P(\hat{p} < 0.01) = P\left( z < \frac{0.01 - 0.02}{0.0062} \right) = P(z < -1.61) = 0.0537 \]
5.37%

8.4 Sampling distribution of a sample mean

The sampling distribution of a sample mean is the distribution of values of the sample mean in all possible random samples of the same size n taken from the same population.

\[ \bar{X} \quad \text{sample mean} \]
What do we expect of sample mean $\bar{X}$?

1. The values of $\bar{X}$ vary from sample to sample in a predictable way.

2. The center of the distribution of $\bar{X}$ is at the true mean $\mu$ (for any sample size $n$).

3. With a larger sample size $n$, the $\bar{X}$ values tend to be closer to the population mean $\mu$.

4. When the sample size $n$ is large, the sample mean takes on many possible values, so $\bar{X}$ can be viewed as a continuous random variable with a density curve as its model.