Sec “ 8 ”

Central Limit Theorem

Sampling Distribution of the

- Sample Proportion.
- Difference between two sample proportions.

Sampling Distribution of Sample Proportion

- If X is a **binomial distribution** with parameters $\boldsymbol{n}, \boldsymbol{p}$, given a population with **mean $\boldsymbol{np}$ and a standard deviation of $\sqrt{\boldsymbol{npq}}$**, **the sampling distribution of the sample proportion $\hat{\boldsymbol{p}}$ has a mean of $\boldsymbol{p}$ and a standard deviation $\sqrt{\frac{\boldsymbol{pq}}{\boldsymbol{n}}}$.**

- We know that $\hat{\boldsymbol{p}} = \frac{\boldsymbol{X}}{\boldsymbol{n}}$

- **The mean of $\hat{\boldsymbol{p}}$:**

i.e., $E(ax) = aE(x)$

$$\mu_{\hat{p}} = E(\hat{p}) = E\left(\frac{X}{n}\right) = \frac{1}{n}E(X) = \frac{1}{n}(np) = p$$

np

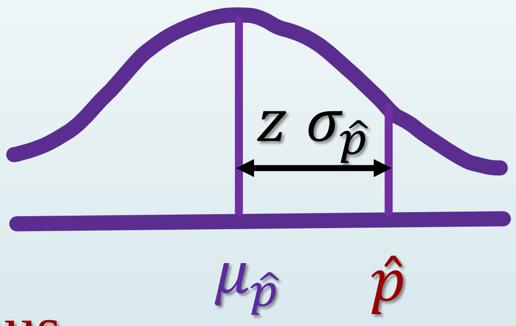
$$\therefore \boldsymbol{\mu_{\hat{p}} = p}$$

➨ **The variance of $\hat{p}$:**

$$\sigma_{\hat{p}}^2 = Var(\hat{p}) = Var\left(\frac{X}{n}\right) = \frac{1}{n^2}Var(X) = \frac{1}{n^2}(npq) = \frac{pq}{n}$$

$\therefore \sigma_{\hat{p}}^2 = \frac{pq}{n}$ → $\therefore \sigma_{\hat{p}} = \sqrt{\frac{pq}{n}}$, where $q = 1 - p$.

Central Limit Theorem



As n increase, the sampling distribution of sample proportion approaches a normal distribution with $\mu_{\hat{p}} = p$, and $\sigma = \sqrt{\frac{pq}{n}}$, $\hat{p} \sim N\left(p, \sqrt{\frac{pq}{n}}\right)$.

The sampling distribution of a sample proportion $\hat{p}$ is approximately normal **<u>if np ≥ 15 and nq ≥ 15.</u>**

Thus,

$\hat{p} = \mu_{\hat{p}} + z\ \sigma_{\hat{p}}$, and

Standard Score

$$z = \frac{\hat{p} - \mu_{\hat{p}}}{\sigma_{\hat{p}}} = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}}$$

Exercise

(7) Your mail-order company advertises that it ships 90% of its orders within three working days. You select a sample random sample of 100 of the 5000 orders received in the past week for an audit.

$$p = 0.9, \quad q = 1 - p = 0.1, \quad n = 100$$

a. What is the mean of the sampling distribution of $\hat{p}$?

Solution

$$\mu_{\hat{p}} = p = 0.9$$

b. Find the standard deviation of the sampling distribution of $\hat{p}$?

Solution

$$\sigma_{\hat{p}} = \sqrt{\frac{pq}{n}} = \sqrt{\frac{(0.9)(0.1)}{100}} = 0.03$$

c. Is the sampling distribution approximately Normal? Check that the Normal conditions are met.

<u>**Solution**</u>

- The expected number of successes must be at least 10

$$np = (100)(0.9) = 90 \checkmark \geq 15$$

- The expected number of failure must be at least 10

$$nq = (100)(0.1) = 10 \times < 15$$

$\therefore$ The distribution of $\hat{p}$ is not approximately normal distribution.

d. **The audit reveals that 86 of theses orders were shipped on time (that is 86%). What is the probability that the proportion of on time orders is 86% or less?**

<u>Solution</u>

$$\mathrm{P}(\hat{p} \leq 0.86) = \mathrm{P}\left(\frac{\hat{p} - \mu_{\hat{p}}}{\sigma_{\hat{p}}} \leq \frac{0.86 - 0.9}{0.03}\right)$$

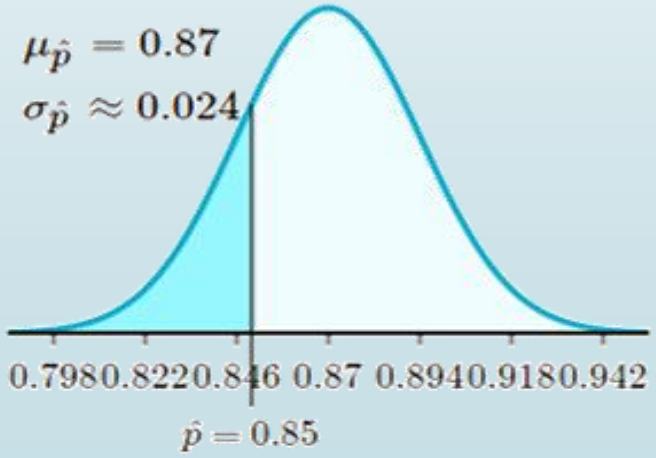
$$= \mathrm{P}(z \leq -1.3333) = \mathrm{P}(z \leq -1.33) = 0.0918 \text{ (from z-table)}$$

(8) According to the US Census Bureau's American Community Survey, 87% of Americans over the age of 25 have earned a high school diploma. Suppose we are going to take a random sample of 200 Americans in this age group and calculate what proportion of the sample has a high school diploma. What is the probability that the proportion of people in the sample with a high school diploma is less than 85%?

Solution

$$p = 0.87, \qquad q = 1 - p = 0.13, \qquad n = 200$$

Note that: The sampling distribution of a sample proportion $\hat{p}$ is approximately normal as long as **the expected number of successes and failures are both at least 15. thus, $np = (200)(0.87) = 174$ and $nq = (200)(0.13) = 26$. So, the distribution of $\hat{p}$ is approximately normal** with $\mu_{\hat{p}} = p = 0.87$, $\sigma_{\hat{p}} = \sqrt{\frac{pq}{n}} = \sqrt{\frac{(0.87)(0.13)}{200}} = 0.0238$



$$P(\hat{p} < 0.85) = P\left(\frac{\hat{p} - \mu_{\hat{p}}}{\sigma_{\hat{p}}} < \frac{0.85 - 0.87}{0.0238}\right)$$

$$= P(z < -0.8403)$$

$$= P(z < -0.84) = 0.2005 \text{ (from z-table)}$$

(9) A USA Today Poll asked a random sample of 1012 U.S. adults what they do with the milk in the bowl after they have eaten the cereal. Of the respondents, 70% said that they drink it. Let $\hat{p}$ be the proportion of people in the sample who drink the cereal milk.

a. **What is the mean of the sampling distribution of $\hat{p}$?**
b. **Find the standard deviation of the sampling distribution of $\hat{p}$?**
c. **Is the sampling distribution approximately Normal? Check that the Normal conditions are met.**
d. **Find the probability of obtaining a sample of 1012 adults in which 67% or fewer say they drink the cereal milk.**

Sampling Distribution of Difference Between Two Proportions

- Suppose we have **two** **populations** with proportions equal to $\boldsymbol{p_1}$ **and** $\boldsymbol{p_2}$. Suppose further that we take all possible samples of size n_1 and n_2 , then the **sampling distribution of difference between proportions** follows a normal distribution with mean $\mu_{\hat{p}_1-\hat{p}_2} = p_1 - p_2$ and standard deviate $\sigma_{\hat{p}_1-\hat{p}_2} = \sqrt{\frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}}$.
- $\mu_{\hat{p}_1-\hat{p}_2} = E(\hat{p}_1 - \hat{p}_2) = E(\hat{p}_1) - E(\hat{p}_2) = \mu_{\hat{p}_1} - \mu_{\hat{p}_2}$

We know that $\mu_{\hat{p}} = p$, therefore $\mu_{\hat{p}_1} = p_1$ and $\mu_{\hat{p}_2} = p_2$.

$$\therefore \mu_{\hat{p}_1-\hat{p}_2} = p_1 - p_2$$

$$\sigma^2_{\hat{p}_1-\hat{p}_2} = Var(\hat{p}_1 - \hat{p}_2) = Var(\hat{p}_1) + Var(-\hat{p}_2)$$
$$= Var(\hat{p}_1) + Var(\hat{p}_2) = \sigma^2_{\hat{p}_1} + \sigma^2_{\hat{p}_2}$$

We know that $\sigma^2_{\hat{p}} = \frac{pq}{n}$, therefore $\sigma^2_{\hat{p}_1} = \frac{p_1 q_1}{n_1}$ and $\sigma^2_{\hat{p}_2} = \frac{p_2 q_2}{n_2}$.

$$\therefore \sigma^2_{\hat{p}_1-\hat{p}_2} = \frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2} \longrightarrow \therefore \sigma_{\hat{p}_1-\hat{p}_2} = \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$$

- The sampling distribution of the difference between two proportions $\hat{p}_1 - \hat{p}_2$ is approximately normal if :

$$n_1 p_1 \geq 15 \text{ and } n_1 q_1 \geq 15,$$

also

$$n_2 p_2 \geq 15 \text{ and } n_2 q_2 \geq 15,$$

Expected successes and failures in both samples at least 15.

- **The standard normal distribution of the difference between two proportions comes by the expression:**

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - \mu_{\hat{p}_1 - \hat{p}_2}}{\sigma_{\hat{p}_1 - \hat{p}_2}} = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}}$$

Exercise:

(10) In one state, 52% of the voters are Republicans, and 48% are Democrats. In a second state, 47% of the voters are Republicans, and 53% are Democrats. Suppose 100 voters are surveyed from each state. Assume the survey uses simple random sampling. What is the probability that the survey will show a greater percentage of Republican voters in the first state than in the second state?

Solution

1^{st} state	2^{nd} state
$p_1 = 0.52$	$p_2 = 0.47$
$q_1 = 0.48$	$q_2 = 0.53$
$n_1 = 100$	$n_2 = 100$

- Make sure the samples from each population are big enough to model differences with a normal distribution:

$$n_1 p_1 = (100)(0.52) = 52 \quad \surd \geq 15$$
$$n_1 q_1 = (100)(0.48) = 48 \quad \surd \geq 15$$

and

$$n_2 p_2 = (100)(0.47) = 47 \quad \surd \geq 15$$
$$n_2 q_2 = (100)(0.53) = 53 \quad \surd \geq 15$$

- This problem requires us to find the probability that $\hat{p}_1$ is less than $\hat{p}_2$. Thus,

$$P(\hat{p}_1 > \hat{p}_2) = P(\hat{p}_1 - \hat{p}_2 > 0)$$

$$= P\left(\frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}} > \frac{0 - (0.52 - 0.47)}{\sqrt{\frac{(0.52)(0.48)}{100} + \frac{(0.47)(0.53)}{100}}}\right)$$

$$= P\left(z > \frac{-0.05}{0.07062}\right) = p(z > -0.7080)$$

$$= p(z > -0.71) = 1 - P(z < -0.71)$$

$$= 1 - 0.2389 = 0.7611$$

		Sampling distribution		
Parameter	**statistic**	**Center(mean)**	**Spread(standard deviation"error")**	**Z-Score**
• P= population proportion.	$\hat{p} =$ sample proportion	$\mu_{\hat{p}} = p$	$\sigma_{\hat{p}} = \sqrt{\frac{pq}{n}}$	$Z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}}$ If $np \geq 15$ and $nq \geq 15$
• $p_1, p_2 =$ the proportion of population1,2	$\hat{p}_1 - \hat{p}_2$	$\mu_{\hat{p}_1-\hat{p}_2} = p_1 - p_2$	$\sigma_{\hat{p}_1-\hat{p}_2} = \sqrt{\frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}}$	$Z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}}}$ If $n_1p_1 \geq 15$ and $n_1q_1 \geq 15$, $n_2p_2 \geq 15$ and $n_2q_2 \geq 15$