Chapter Six
Introduction to Sampling Distributions

Chapter Goals:

After completing this chapter, you should be able to:

- Define the concept of sampling error.
- Determine the mean and standard deviation for the sampling distribution of the sample mean $\bar{X}$.
- Determine the mean and standard deviation for the sampling distribution of the sample proportion $\bar{p}$.
- Describe the Central Limit Theorem and its importance.
- Apply sampling distributions for both $\bar{X}$ and $\bar{p}$.

6.1. Sampling Error

Sample Statistics are used to estimate Population Parameters;

The sample mean: $\bar{X} = \frac{\sum x}{n}$ is an estimate of the population mean, μ.

The sample standard deviation: $S = \sqrt{\frac{\sum x^2 - n\bar{x}^2}{n-1}}$ is an estimate of the population standard deviation, σ.

Note:

Different samples provide different estimates of the population parameter. Sample results have potential variability, thus sampling error exits.

Sampling error:

The difference between a value computed from a sample (a statistic) and the corresponding value computed from a population (a parameter).

$$\text{Sampling Error} = \bar{X} - \mu$$

Where:

$$\bar{X} = \frac{\sum x}{n} \text{ sample mean} \qquad \text{and} \qquad \mu = \frac{\sum X}{N} \text{ population mean}$$

Example: If the population mean is μ = 99.2 degrees and a sample of n = 5 temperatures yields a sample mean of $\bar{x}$ = 98.6 degrees, then the sampling error is $\bar{X} - \mu = 98.6 - 99.2 = -0.6$ degrees.

Notes:

- Different samples will yield different sampling errors.
- The sampling error may be positive or negative ($\bar{x}$ may be greater than or less than μ).
- The expected sampling error decreases as the sample size increases.

Example: Q.6.1 Page 233

6.2. Sampling Distribution

A sampling distribution is a distribution of the possible values of a statistic for a given sample size selected from a population.

Developing a Sampling Distribution

Assume there is a population, the size of the population is N = 4 (A, B, C, D), Let *X* be a random variable that defined the age of individuals

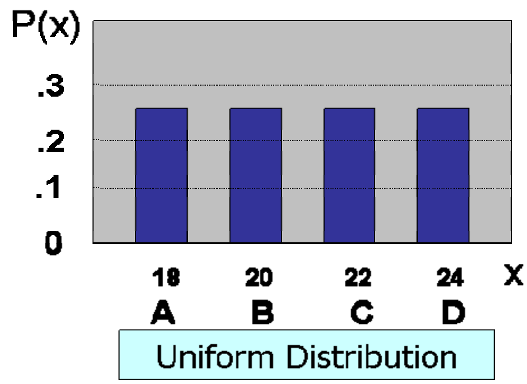
Values of *x*: 18 for A, 20 for B, 22 for C, 24 for D (years).

Then

Summary Measures for the Population Distribution:

$$\mu = \frac{\sum x_i}{N} = \frac{18+20+22+24}{4} = 21$$

$$\sigma = \sqrt{\frac{\sum (x_i - \mu)^2}{N}} = 2.236$$



Now

Select samples each sample of size 2, then find the sample mean for each sample.

There is 16 possible samples (sampling with replacement)

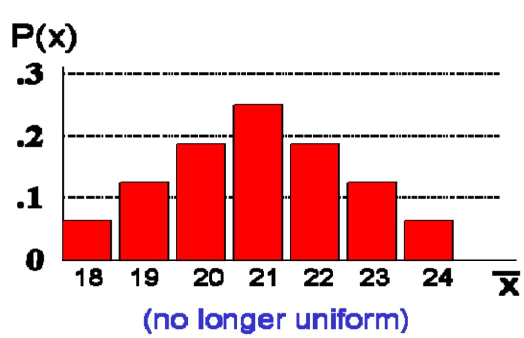
1st observation	2nd observation			
	18	**20**	**22**	**24**
18	(18,18)	(18,20)	(18,22)	(18,24)
20	(20,18)	(20,20)	(20,22)	(20,24)
22	(22,18)	(22,20)	(22,22)	(22,24)
24	(24,18)	(24,20)	(24,22)	(24,24)

There is 16 Sample Means

1st observation	2nd observation			
	18	**20**	**22**	**24**
18	18	19	20	21
20	19	20	21	22
22	20	21	22	23
24	21	22	23	24

The sampling distribution for all the sample means is not uniform, it has bell shape

and



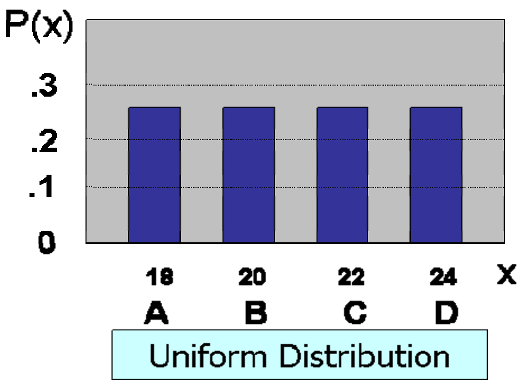
$$\mu_{\bar{X}} = \frac{\sum \bar{X}_i}{N} = \frac{18+19+21+\cdots+24}{16} = 21$$

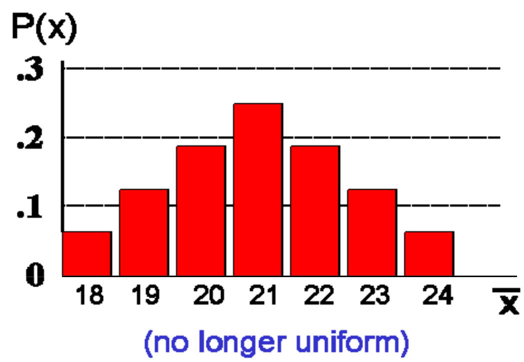
$$\sigma_{\bar{X}} = \sqrt{\frac{\sum (X_i - \mu_{\bar{X}})^2}{N}}$$

$$= \sqrt{\frac{(18-21)^2 + (19-21)^2 + \cdots + (24-21)^2}{16}} = 1.58$$

Comparing the Population with its Sampling Distribution

Population **Sample**





$N=4,\quad \mu = 21,\quad \sigma = 2.236$ $n=2,\quad \mu_{\bar{x}} = 21,\quad \sigma_{\bar{x}} = 1.58$

Theorem:

If a population is *normal* with mean μ and standard deviation σ, the sampling distribution of $\bar{X}$ is also normal with

Mean $\mu_{\bar{X}} = \mu$ and standard deviation $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$

Note:

For normal population distributions, the sampling distribution of the mean is always normally distributed

Z -value for the sampling distribution of $\bar{X}$:

$$Z = \frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}} = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

where: $\bar{X}$ = sample mean, μ = population mean, σ = population standard deviation n = sample size.

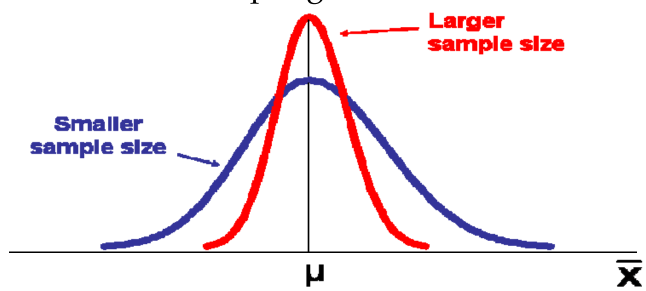
Apply the Finite Population Correction (CF):

If the sample is large relative to the population (n is greater than 5% of N) and the Sampling is without replacement, then

$$Z = \frac{(\overline{X} - \mu)}{\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}}$$

Sampling Distribution Properties

1. The Normal Population Distribution and Normal Sampling Distribution has the same mean $\mu_{\overline{X}} = \mu$.
2. For sampling with replacement: As n increases, $\sigma_{\overline{X}}$ decreases.



Example: A sample of size 25 is selected randomly from a normal distribution with a mean 70 and a standard deviation 16, Find;

a. The sampling distribution for the mean.

b. The probability that the sample mean is grater than 73.

c. The median of the sample mean.

Example: Q 6.14 page 248:

If the Population is not Normal

The Central Limit Theorem (CLT):

Even if the population is not normal, the sample means from the population will be *approximately* normal as long as the sample size is large enough $(n \geq 30)$ and the sampling distribution will have

$$\mu_{\bar{x}} = \mu \text{ and } \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

Note:

- For most distributions, $n \geq 30$ will give a sampling distribution that is nearly normal.
- For fairly symmetric distributions, $n \geq 15$.

Example: A random sample of size 100 is taken from an infinite population having a mean 76 and a variance 256. Find;

a. The sampling distribution for the sample mean.

b. What is the probability that the sample mean will be between 75 and 78.

Example: Suppose a population has mean μ = 8 and standard deviation σ = 3. Suppose a random sample of size n = 36 is selected.

a. Find the sampling distributionof the sample mean.

b. What is the probability that the sample mean is between 7.8 and 8.2?

Example: Suppose a population has an exponential distribution with mean 10. If a sample of size 100 was selected,

a. Find the sampling distribution of the sample mean.

b. What is the probability that the sample mean will exceed 10.5.

6.3: Sampling Distribution for the Proportion:

Population proportion: the ratio of number of items having a cretain attribute in the population to the size of the population, where $p = \frac{X}{N}$.

Sample proportion: the ratio of number of items having a certain attribute in the sample to the size of the sample, where $\overline{p} = \frac{x}{n}$ and the sampling error = $\overline{p} - p$.

Note: If there are two outcomes, then $\overline{p}$ has a binomial distribution.

Sampling distribution for $\overline{p}$ for large sample size ($n \ge 30$)

If $np \ge 5$ and $nq = n(1-p) \ge 5$, then $\overline{p}$ has *approximately normal* distribution with mean $\mu_{\overline{p}} = p$ and standard deviation $\sigma_{\overline{p}} = \sqrt{\frac{p(1-p)}{n}}$

Standardize *p* to a *Z* value with the formula:

$$Z = \frac{\overline{p} - p}{\sigma_p} = \frac{\overline{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$$

Note: If sampling is **without replacement** and n is greater than 5% of the population size, then σ_p must use the **finite population correction factor:**

$$\sigma_p = \sqrt{\frac{p(1-p)}{n}}\sqrt{\frac{N-n}{N-1}}$$

Example: If the true proportion of voters who support Proposition A is $p = 0.4$, what is the probability that a sample of size 200 yields a proportion, of voters with YES, between 0.40 and 0.45?
i.e.: if $p = 0.4$ and $n = 200$, what is $P(0.40 \le \overline{p} \le 0.45)$?

Solution:

$np = (200)(0.4) = 80 \ge 5$, and $nq = (200)(0.6) = 120 \ge 5$.

then $\overline{p}$ has approximately normal distribution with mean $\mu_{\overline{p}} = p = 0.4$ and standard deviation $\sigma_{\overline{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.4(1-0.4)}{200}} = 0.03464$

$$P(0.40 \le \overline{p} \le 0.45) = P\left(\frac{0.40-0.40}{0.03464} \le Z \le \frac{0.45-0.40}{0.03464}\right) = P(0 \le Z \le 1.44) = \Phi(1.44) = 0.4251$$

Example: Q6.32 Page 256

$p = 0.5 \quad , n = 200$

$np = 0.5(200) = 100 \ge 5$ and $nq = n(1-p) = 200(0.5) = 100 \ge 5$

then $\bar{p}$ has approximately normal distribution with

mean $\mu_{\bar{p}} = p = 0.5$

standard deviation $\sigma_{\bar{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.5(0.5)}{200}} = 0.0354$

$$P(0.47 < \bar{p} < 0.51) = P\left(\frac{0.47-0.5}{0.0354} < Z < \frac{0.51-0.5}{0.0354}\right) = P(-0/35 < Z < 0.28)$$
$$= 0.1103 + 0.3023 = 0.4126$$

Example: Q6.38 Page 256

a. $p = 0.1 \quad , n = 100,$ and $x = 14 \Rightarrow \bar{p} = \frac{x}{n} = \frac{14}{100} = 0.14$

$np = 100(0.1) = 10 \geq 5$ and $nq = n(1-p) = 100(0.9) = 90 \geq 5$

$\bar{p}$ has approximately normal distribution with

mean $\mu_{\bar{p}} = p = 0.1$

standard deviation $\sigma_{\bar{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.1(0.9)}{100}} = 0.03$

$$P(\bar{p} > 0.14) = P\left(Z > \frac{0.14-0.1}{0.03}\right) = P(Z > 1.33) = 0.5 - 0.4082 = 0.0918$$

b. $p = 0.1 \quad , n = 100 + 100 = 200, \ \textit{and} \ \ x = 14 + 14 = 28 \Rightarrow \bar{p} = \frac{28}{200} = \frac{14}{100} = 0.14$

$np = (200)0.1 = 20 \geq 5$ and $nq = n(1-p) = (200)0.9 = 180 \geq 5$

$\bar{P}$ has approximately normal distribution with

mean $\mu_{\bar{p}} = p = 0.1$

standard deviation $\sigma_{\bar{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.1(0.9)}{200}} = 0.021213203$

$$P(\bar{p} > 0.14) = P\left(Z > \frac{0.14-0.1}{0.021213203}\right) = P(Z > 1.89) = 0.5 - 0.4706 = 0.0249$$

Both a and b support the VIP claim