Sampling, Sampling Distribution and Normality

Presented by:
Mahendra Adhi Nugroho, M.Sc

Sources:
Anderson, Sweeney,wiliams, ***Statistics for Business and Economics*** , 6 e, Pearson education inc, 2007
Sugiyono, ***Statistika untuk penelitian***, alfbeta, Bandung, 2007

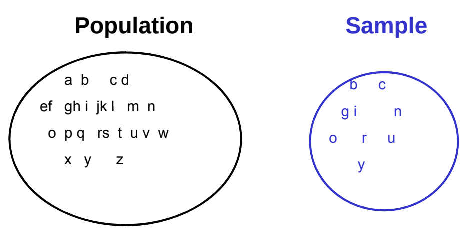
Tools of Business Statistics

- **Descriptive statistics**
  - Collecting, presenting, and describing data
- **Inferential statistics**
  - Drawing conclusions and/or making decisions concerning a population based only on sample data

Populations and Samples

- A **Population** is the set of all items or individuals of interest
  - Examples: All likely voters in the next election
    All parts produced today
    All sales receipts for November
- A **Sample** is a subset of the population
  - Examples: 1000 voters selected at random for interview
    A few parts selected for destructive testing
    Random receipts selected for audit

Population vs. Sample



Why Sample?

- Less time consuming than a census
- Less costly to administer than a census
- It is possible to obtain statistical results of a sufficiently high precision based on samples.

Sampling Methods

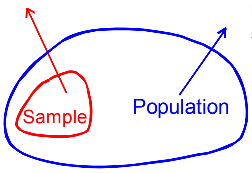
Sampling Methods

- Probability sampling
  1. Simple random sampling
  2. Proportionate stratified random sampling
  3. Disproportionate stratified random sampling
  4. Cluster sampling
- Non probability sampling
  1. Systematic sampling
  2. Quota sampling
  3. Incidental sampling
  4. Purposive sampling
  5. Surfeited sampling
  6. Snowball sampling

Inferential Statistics

- Making statements about a population by examining sample results

Sample statistics → Population parameters
(known) **Inference** (unknown, but can be estimated from sample evidence)



Inferential Statistics

Drawing conclusions and/or making decisions concerning a population based on sample results.

- **Estimation**
  - e.g., Estimate the population mean weight using the sample mean weight
- **Hypothesis Testing**
  - e.g., Use sample evidence to test the claim that the population mean weight is 120 pounds

Sampling Distributions

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

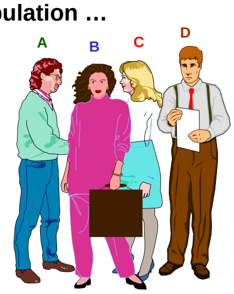
Sampling Distributions of Sample Means

Sampling Distributions
- Sampling Distribution of Sample Mean
- Sampling Distribution of Sample Proportion
- Sampling Distribution of Sample Variance

Note: this chapter only discussing sample mean distribution.

Developing a Sampling Distribution

- **Assume there is a population …**
- Population size N=4
- Random variable, X, is age of individuals
- Values of X: 18, 20, 22, 24 (years)



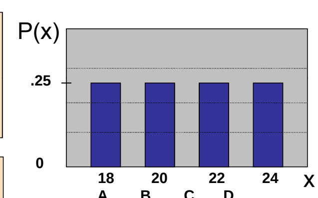
Developing a Sampling Distribution

(continued)

Summary Measures for the Population Distribution:

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

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



Uniform Distribution

Developing a Sampling Distribution

(continued)

Now consider all possible samples of size n = 2

1st Obs	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

16 possible samples (sampling with replacement)

16 Sample Means

1st Obs	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

Developing a Sampling Distribution

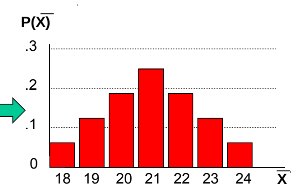
(continued)

Sampling Distribution of All Sample Means

16 Sample Means

1st Obs	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

Sample Means Distribution



(no longer uniform)

Developing a Sampling Distribution

(continued)

Summary Measures of this Sampling Distribution:

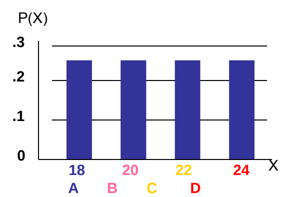
$$E(\bar{X}) = \frac{\sum \bar{X}_i}{N} = \frac{18+19+21+\cdots+24}{16} = 21 = \mu$$

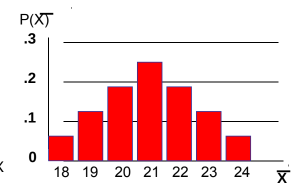
$$\sigma_{\bar{X}} = \sqrt{\frac{\sum(\bar{X}_i - \mu)^2}{N}} = \sqrt{\frac{(18-21)^2 + (19-21)^2 + \cdots + (24-21)^2}{16}} = 1.58$$

Comparing the Population with its Sampling Distribution

Population
N = 4
$\mu = 21 \quad \sigma = 2.236$

Sample Means Distribution
n = 2
$\mu_{\bar{X}} = 21 \quad \sigma_{\bar{X}} = 1.58$





Expected Value of Sample Mean

- Let $X_1, X_2, \ldots X_n$ represent a random sample from a population
- The sample mean value of these observations is defined as

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Standard Error of the Mean

- Different samples of the same size from the same population will yield different sample means
- A measure of the variability in the mean from sample to sample is given by the Standard Error of the Mean:

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

- Note that the standard error of the mean decreases as the sample size increases

If the Population is Normal

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

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

Z-value for Sampling Distribution of the Mean

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

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

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

Finite Population Correction

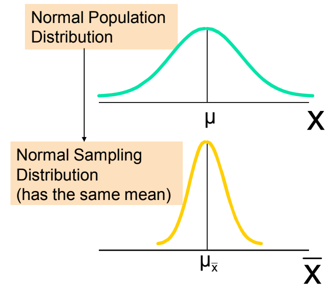
- Apply the **Finite Population Correction** if:
  - a population member cannot be included more than once in a sample (sampling is without replacement), and
  - the sample is large relative to the population (n is greater than about 5% of N)
- Then

$$Var(\bar{X}) = \frac{\sigma^2}{n}\frac{N-n}{N-1} \quad \text{or} \quad \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

Sampling Distribution Properties

- $\mu_{\bar{x}} = \mu$

(i.e. $\bar{X}$ is unbiased)

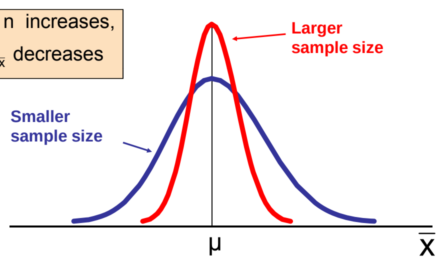


Sampling Distribution Properties

(continued)

- For sampling with replacement:

As n increases, $\sigma_{\bar{x}}$ decreases



If the Population is **not** Normal

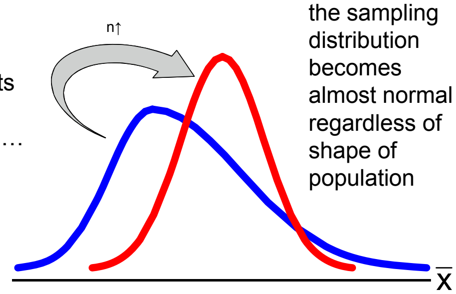
- We can apply the Central Limit Theorem:
  - Even if the population is not normal,
  - …sample means from the population will be approximately normal as long as the sample size is large enough.

Properties of the sampling distribution:

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

Central Limit Theorem

As the sample size gets large enough…



the sampling distribution becomes almost normal regardless of shape of population

If the Population is **not** Normal

(continued)

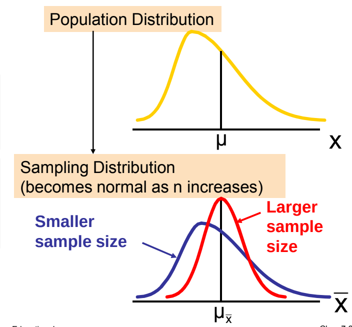
Sampling distribution properties:

Central Tendency

$$\mu_{\bar{x}} = \mu$$

Variation

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$



How Large is Large Enough?

- For most distributions, n > 25 will give a sampling distribution that is nearly normal
- For normal population distributions, the sampling distribution of the mean is always normally distributed

Deciding Sample Size

Isaac and Michael Approach

$$s = \frac{\lambda^2 . N.P.Q}{d^2(N-1)+\lambda^2 .P.Q}$$

- Use table on sugiyono page 71

Nomogram Herry King

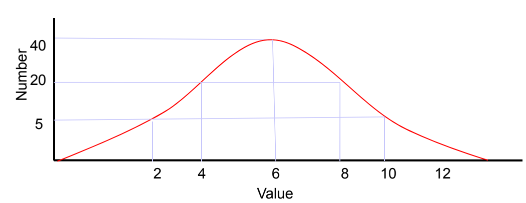
- Maximum sample size is 2000

Suggested Sample Size

- Proper sample size in a research are 30-500 samples
- Proper size in categorized sample are minimum 30 samples each category
- Proper sample size for multivariate data analysis (correlation or multivariate regression) are minimum 10 times of variables numbers (independent ad dependent)
- Proper sample size for simple experimental design that use experiment and control groups are 10 – 20 samples each variable group.

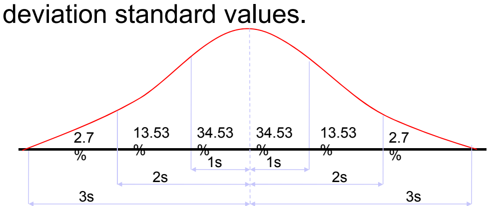
Normality: Normal Curve

- A data set could have normal distribution if sum of data and standard deviation of upper mean and under mean data are same.



Normality: Normal Curve

- Divide normal curve in 6 areas base on deviation standard values.



$$z = \frac{(x_i - \bar{x})}{s}$$

Normality Test Using Chi square

- Decide interval class. In this chase use 6 as class interval because chi square normal distributions is divided in 6 part.
- Decide interval wide
- Counting the estimated chi square value and compare the value with value that is stated in chi square table.

$$\chi^2 = \frac{(f_o - f_e)^2}{f_e}$$

Let me win!
If I cannot be a winner,
Let me brave in attempt!