MATH1. Part II

Probability and Statistics

Chapter 5

Sampling Distributions

5.1 Basic Ideas

Chapter 1.3 Random Sampling

- **Population**. The population consists of all subjects/animals/specimens/plants, and so on, of interest.
- Typically we are unable to observe the entire population; therefore, we must be content with gathering data from a subset of the population, a **sample** of size n. From this sample we make inferences about the population as a whole

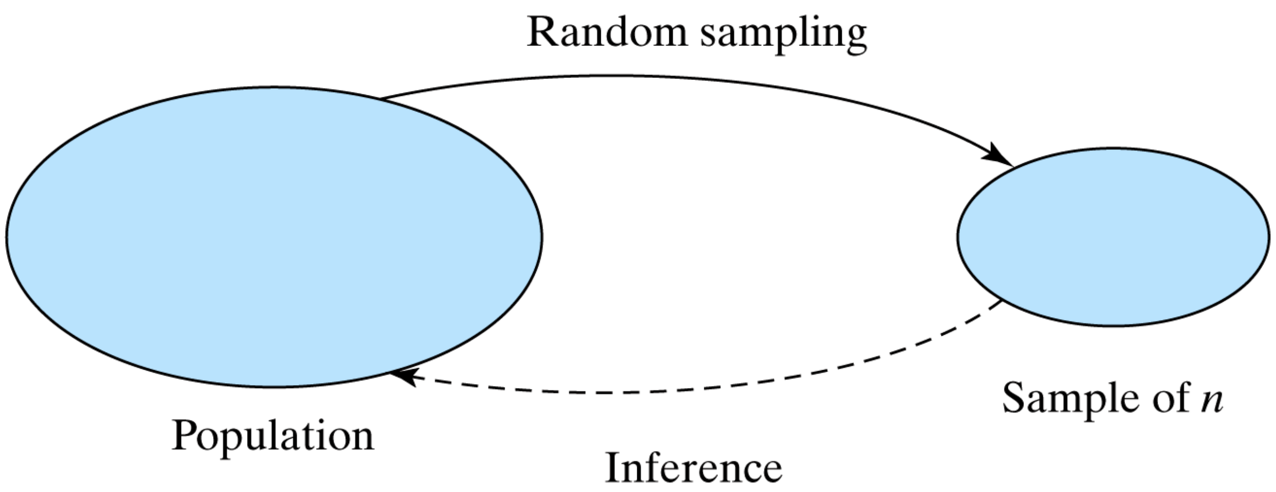


Figure 1.3.1 Sampling from a population

5.1 Basic Ideas

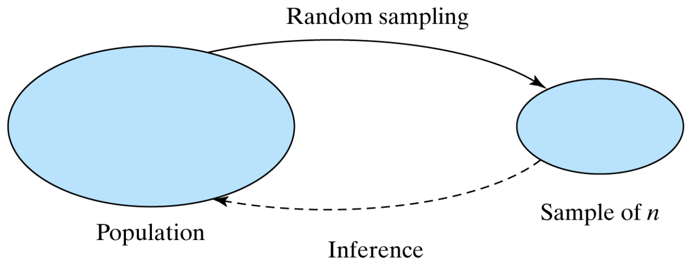
Chapter 1.3 Random Sampling

A Simple Random Sample

- A simple random sample of n items is a sample in which
- (a) every member of the population has the <u>same chance</u> of being included in the sample, and
- (b) the members of the sample are chosen <u>independently</u> of each other. [Requirement (b) means that the chance of a given member of the population being chosen does not depend on which other members are chosen.]

Sampling error

- The discrepancy between the sample and the population is called chance error due to sampling or sampling error.



5.1 Basic Ideas

Sampling Variability

An important goal of data analysis is to distinguish between features of the data that reflect real biological facts and features that may reflect only chance effects.

– The underlying reality (real biological facts) is visualized as a population,

– the data are viewed as a random sample from the population, and

– chance effects are regarded as sampling error—that is, discrepancy between the sample and the population.

- The variability among random samples from the same population is called **sampling variability**.

5.1 Basic Ideas

The Meta-Study

- A **meta-study** consists of indefinitely many repetitions, or replications, of the same study.
  - Thus, if the study consists of drawing a random sample of size n from some population, the corresponding **meta-study** involves drawing repeated random samples of size n from the same population.
- The process of repeated drawing is carried on indefinitely, with the members of each sample being replaced before the next sample is drawn.

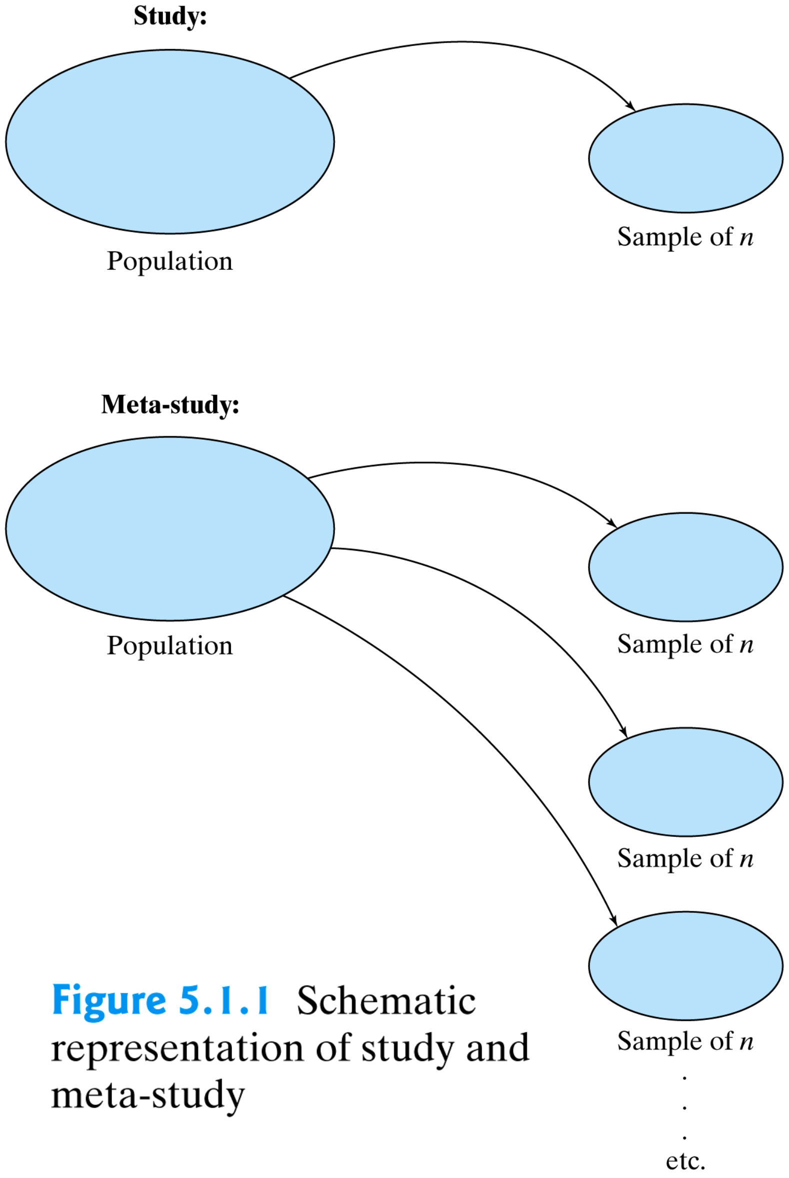


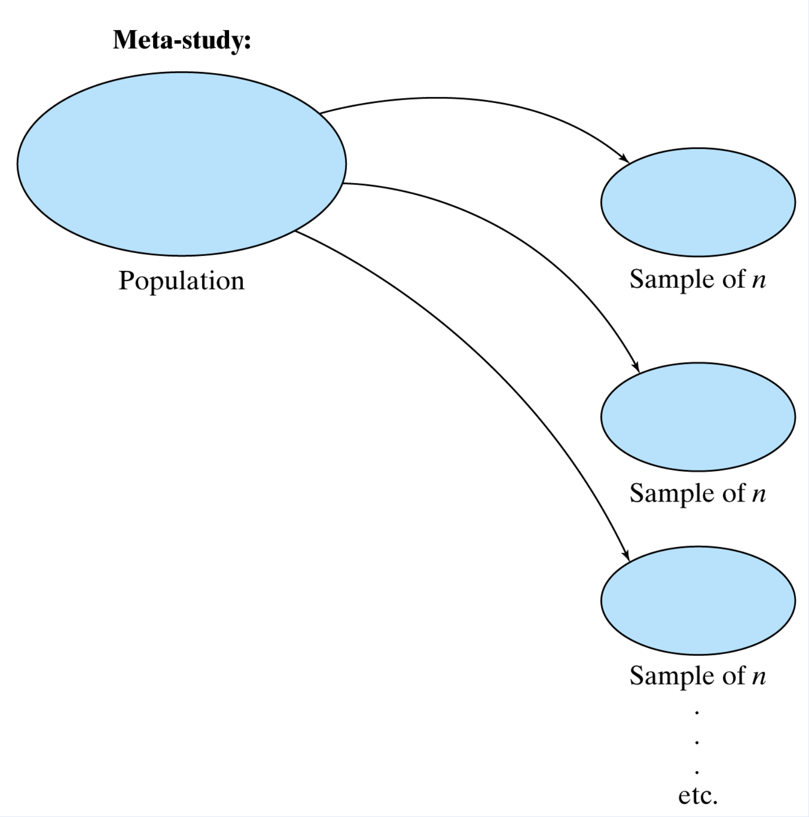
Figure 5.1.1 Schematic representation of study and meta-study

5.1 Basic Ideas

The Meta-Study

Example 5.1.1 Rat blood Pressure

- A study consists of measuring the change in blood pressure in each of n = 10 rats after administering a certain drug.
- What is the corresponding meta-study?



Example 5.1.2 Bacterial Growth

- A study consists of observing bacterial growth in n = 5 petri dishes that have been treated identically.
- What is the corresponding meta-study?

5.1 Basic Ideas

The Meta-Study

Example 5.1.1 Rat blood Pressure

- A study consists of measuring the change in blood pressure in each of n = 10 rats after administering a certain drug.
- What is the corresponding meta-study?
  - The corresponding meta-study would consist of <u>repeatedly</u> choosing groups of n = 10 rats from the same population and making blood pressure measurements under the same conditions.

Example 5.1.2 Bacterial Growth

- A study consists of observing bacterial growth in n = 5 petri dishes that have been treated identically.
- What is the corresponding meta-study?
  - The corresponding meta-study would consist of repeatedly preparing groups of five petri dishes and observing them in the same way.

5.1 Basic Ideas

The Meta-Study

- The meta-study concept provides a link between sampling variability and probability.
- Chapter 3: the probability of an event can be interpreted as the long-run relative frequency of occurrence of the event.
- **Probabilities concerning a random sample** can be interpreted as **relative frequencies in a meta-study**.

- The **sampling distribution** describes the variability, for a chosen statistic, among the many random samples in a meta-study.
- Knowing a sampling distribution allows one to **make probability statements about possible samples**.

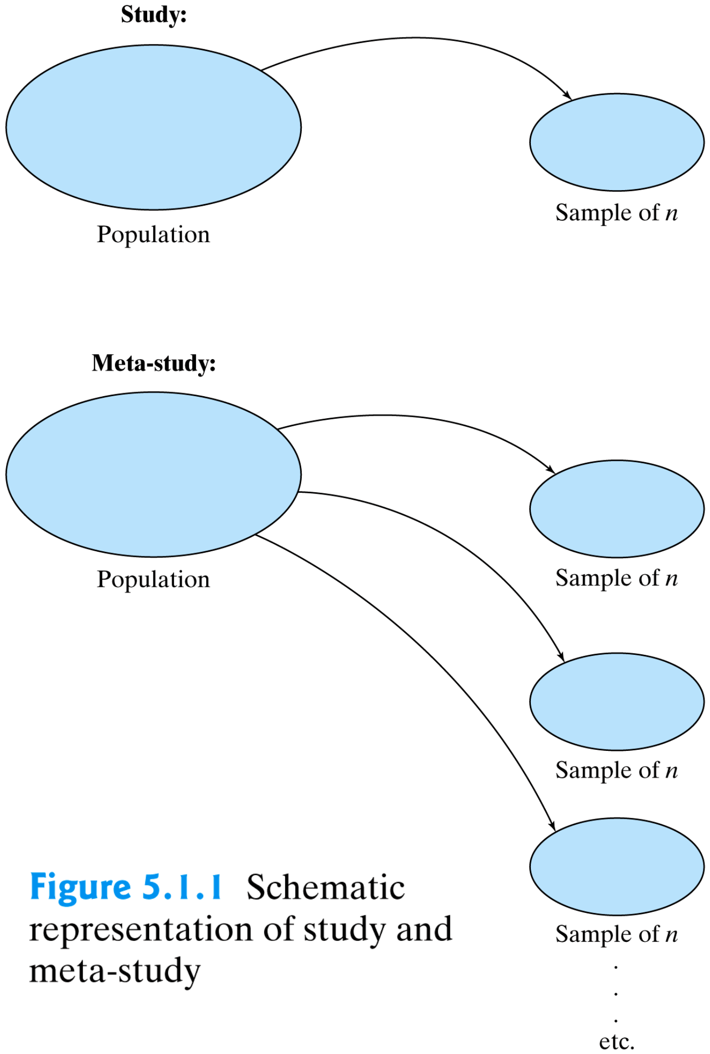
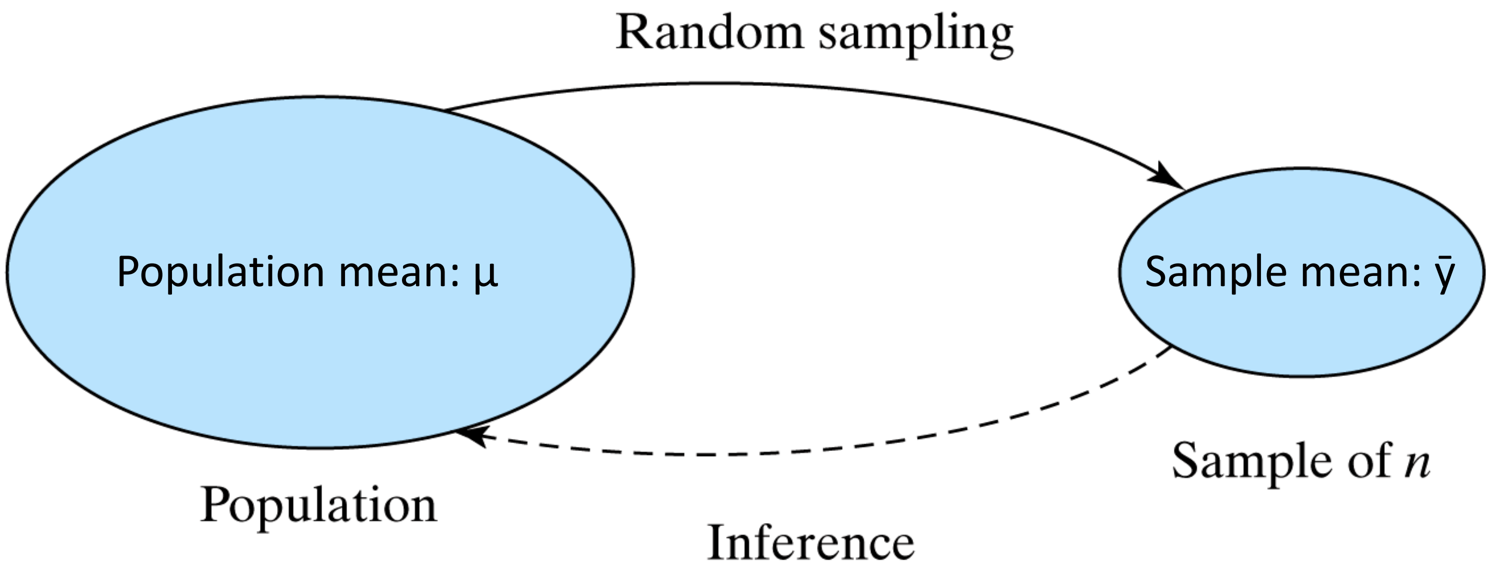


Figure 5.1.1 Schematic representation of study and meta-study

5.2 The Sample Mean

The sampling distribution of $\bar{Y}$

- Sample mean $\bar{y}$ (Chapter 2): the **mean** of a sample (or "the sample mean") is the sum of the observations divided by the number of observations.
- It is natural to ask, **"How close to μ is $\bar{y}$?"**



5.2 The Sample Mean

The sampling distribution of $\bar{Y}$

- if we think in terms of the random sampling model
- Think the **sample mean** as a random variable $\bar{Y}$.
- The question then becomes:

  "How close to μ is $\bar{Y}$ *likely* to be?"

  - the answer is provided by the sampling distribution of $\bar{Y}$
  - that is, the probability distribution that describes sampling variability in $\bar{Y}$.

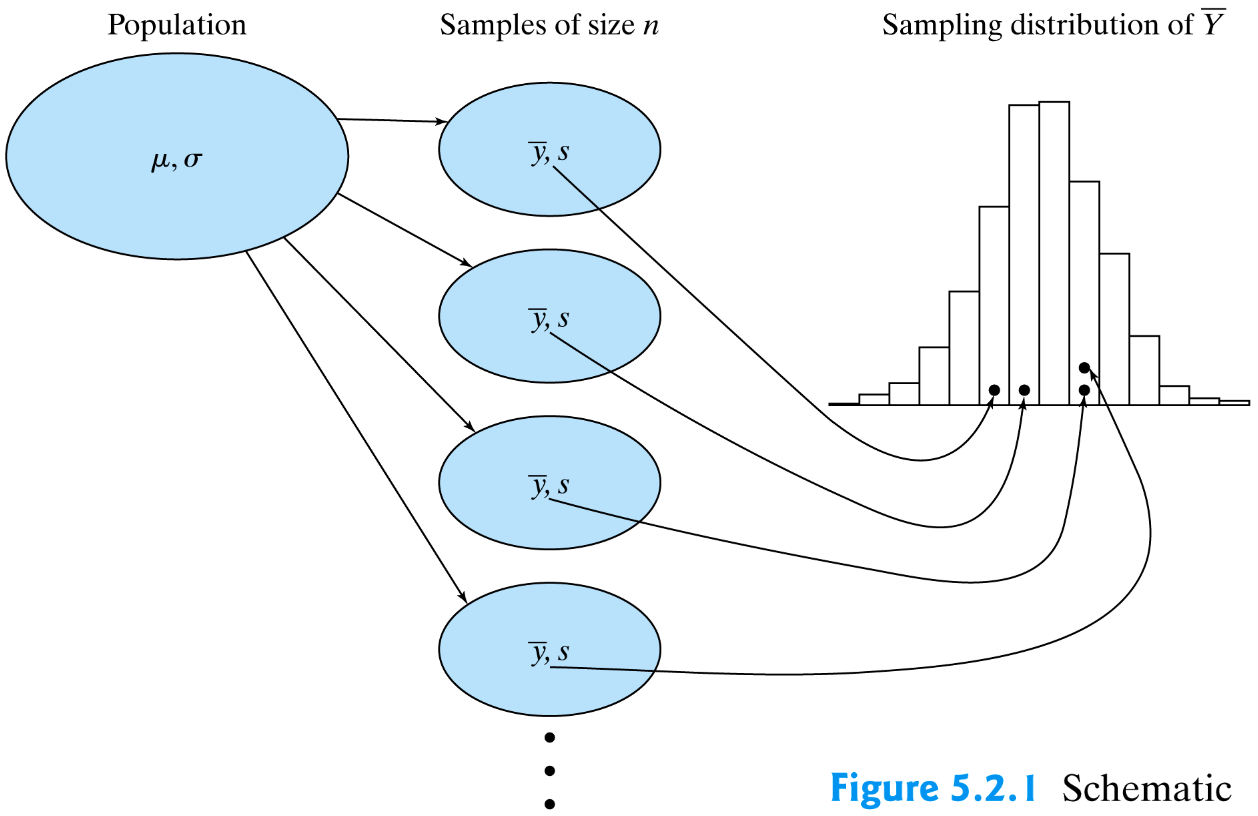


Figure 5.2.1 Schematic representation of the sampling distribution of $\bar{Y}$

5.2 The Sample Mean

Theorem 5.2.1: The Sampling Distribution of $\bar{Y}$

1. **Mean**: The mean of the sampling distribution of $\bar{Y}$ is equal to the population mean.
   — In symbols, $\mu_{\bar{Y}} = \mu$
2. **Standard deviation**: The standard deviation of the sampling distribution of $\bar{Y}$ is equal to the population standard deviation divided by the square root of the sample size.
   — In symbols, $\sigma_{\bar{Y}} = \sigma/\sqrt{n}$
   — As the sample size goes up, the variability in the sample mean goes down.

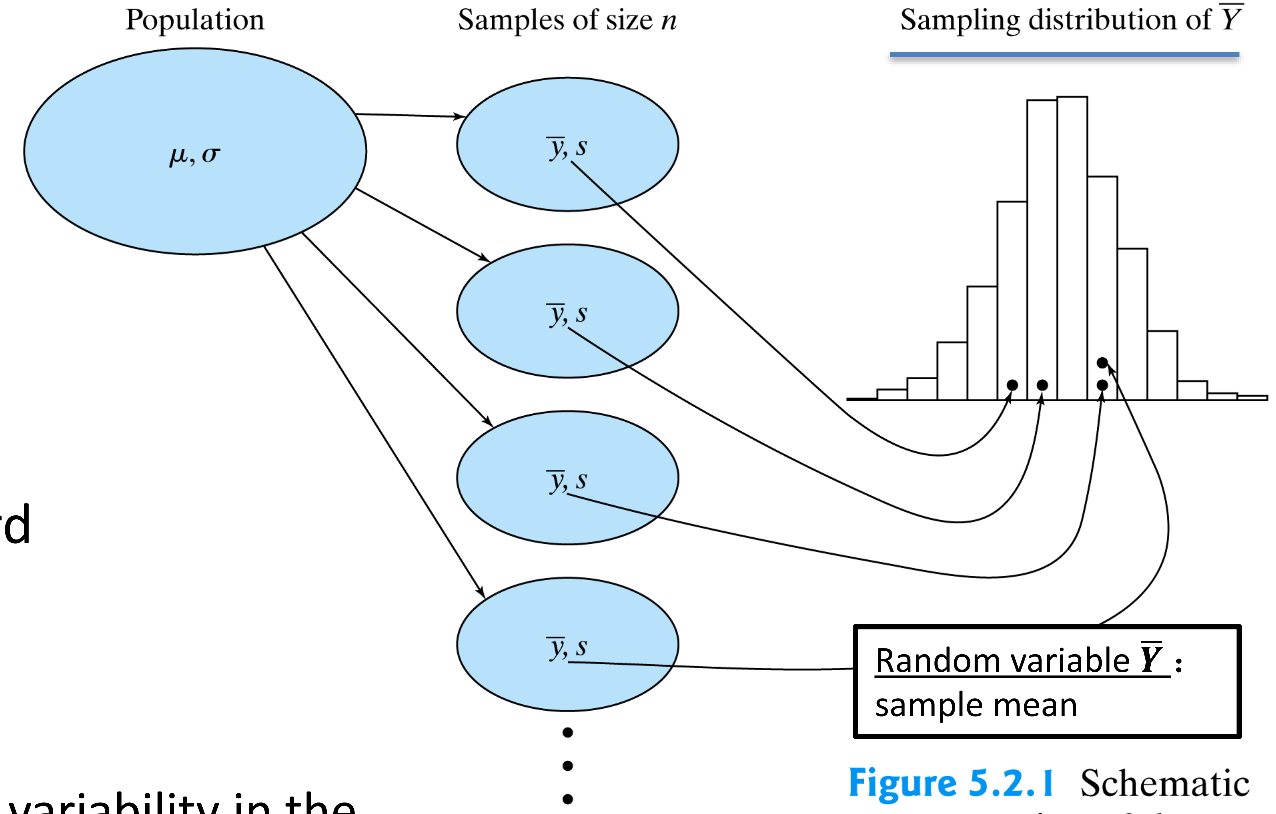


Figure 5.2.1 Schematic representation of the sampling distribution of $\bar{Y}$

5.2 The Sample Mean

Theorem 5.2.1: The Sampling Distribution of $\overline{Y}$

Prove Theorem 5.2.1.

- $\mu_{\bar{Y}} = \mu$
- $\sigma_{\bar{Y}} = \sigma/\sqrt{n}$

Hint: a meta study with sample size n is equivalent to study the whole population for n times.

5.2 The Sample Mean

Theorem 5.2.1: The Sampling Distribution of $\overline{Y}$

Prove Theorem 5.2.1.

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

Proof:

$$\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

$$E(\overline{Y}) = E(\frac{1}{n}\sum_{i=1}^{n} y_i) = \frac{1}{n} E\left(\sum_{i=1}^{n} (y_i)\right) = \frac{1}{n}\sum_{i=1}^{n} E(y_i) = \frac{1}{n} \times n\, \mu = \mu$$

5.2 The Sample Mean

Theorem 5.2.1: The Sampling Distribution of $\bar{Y}$

Prove Theorem 5.2.1.

- $\sigma_{\bar{Y}} = \sigma/\sqrt{n}$

Proof:

$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} y_i$

$\mathrm{Var}(\bar{Y}) = \mathrm{Var}(\frac{1}{n}\sum_{i=1}^{n} y_i) = \frac{1}{n^2}\sum_{i=1}^{n} \mathrm{Var}(y_i) = \frac{1}{n^2} \times n\,\sigma^2 = \frac{\sigma^2}{n}$

$\sigma_{\bar{Y}} = \sigma/\sqrt{n}$

5.2 The Sample Mean

Theorem 5.2.1: The Sampling Distribution of $\overline{Y}$

3. Shape

a) If the population distribution of Y is <u>normal</u>, then the sampling distribution of $\overline{Y}$ is <u>normal</u>, regardless of the sample size n.

b) **Central Limit Theorem**:

If <u>n is large</u>, then the sampling distribution of $\overline{Y}$ is approximately <u>normal</u>, even if the population distribution of Y is not normal.

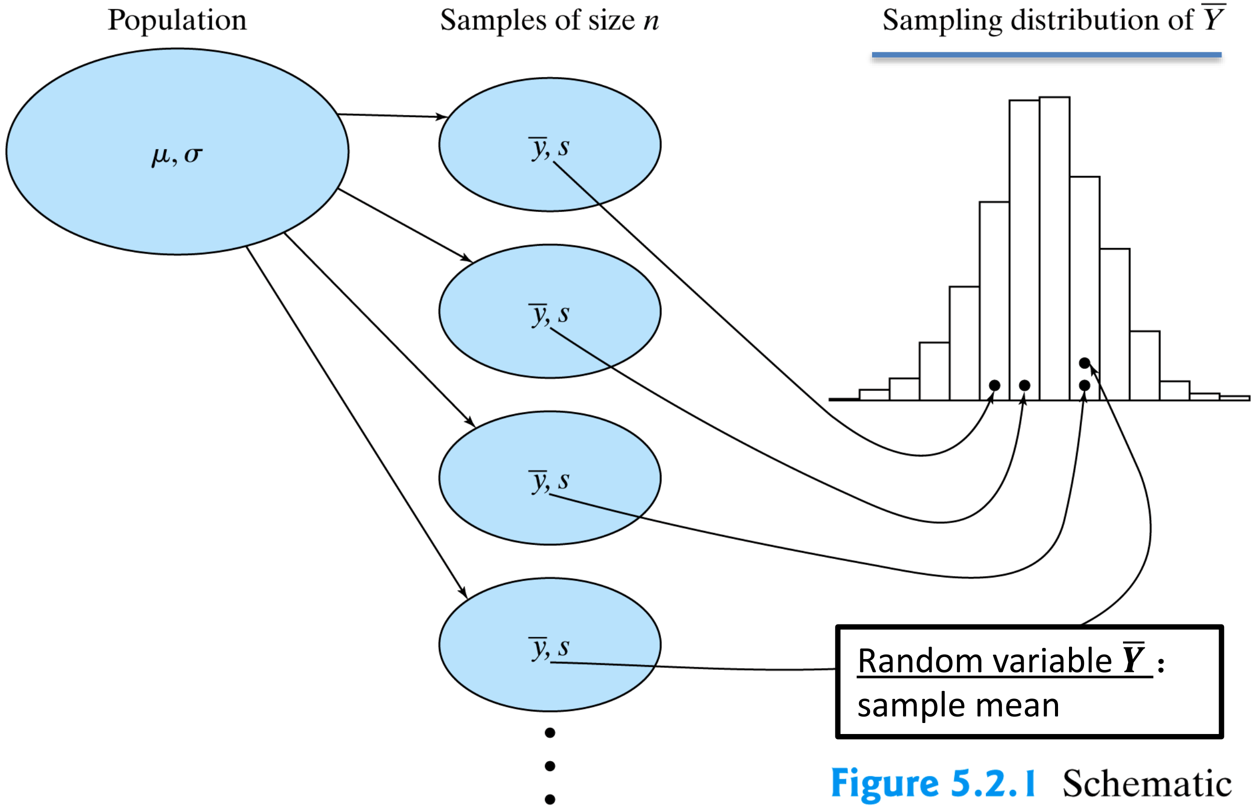


Figure 5.2.1 Schematic representation of the sampling distribution of $\overline{Y}$

5.2 The Sample Mean

Theorem 5.2.1: The Sampling Distribution of $\overline{Y}$ - 3b

Central Limit Theorem: If n is large, then the **sampling distribution of $\overline{Y}$** is approximately **normal**, even if the population distribution of Y is not normal.

- The Central Limit Theorem states that, no matter what distribution Y may have in the population,* if the sample size is large enough, then the sampling distribution of Y will be approximately a normal distribution.
- The Central Limit Theorem is of fundamental importance because it can be applied when (as often happens in practice) the form of the population distribution is not known.
- It is because of the Central Limit Theorem (and other similar theorems) that the normal distribution plays such a central role in statistics.

5.2 The Sample Mean

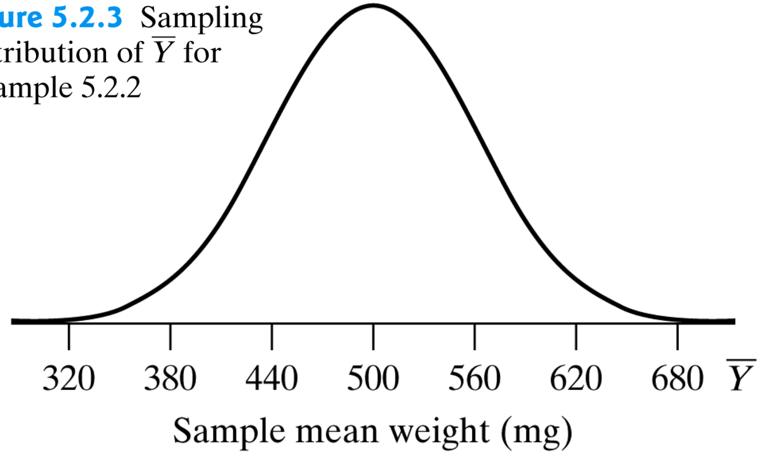
Theorem 5.2.1: The Sampling Distribution of $\bar{Y}$

Example 5.2.2 Weights of seeds

- A large population of seeds of the princess bean *Phaseotus vulgaris* is to be sampled. The weights of the seeds in the population follow a normal distribution with mean **μ = 500** mg and standard deviation **σ = 120** mg.
- Suppose now that a random sample of four seeds is to be weighed, and let $\bar{Y}$ represent the mean weight of the four seeds.

- Use Theorem 5.2.1, what is the mean and standard deviation of sampling distribution of $\bar{Y}$?
- What are the meaning of above values?

Figure 5.2.3 Sampling distribution of $\bar{Y}$ for Example 5.2.2



5.2 The Sample Mean

Theorem 5.2.1: The Sampling Distribution of $\bar{Y}$

Example 5.2.2 Weights of seeds

- the population follow a normal distribution with **μ = 500** mg and **σ = 120** mg; n = 4.
- Use Theorem 5.2.1, what is the mean and standard deviation of sampling distribution of $\bar{Y}$?
- What are the meaning of above values?
  - According to Theorem 5.2.1, the sampling distribution of $\bar{Y}$ will be a normal distribution.
  - $\mu_{\bar{Y}} = \mu = 500$ mg
  - $\sigma_{\bar{Y}} = \sigma/\sqrt{n} = 120/\sqrt{4} = 60$ mg
    - Thus, on average the sample mean will equal 500 mg,
    - but the variability from one sample of size 4 to the next sample of size 4 is such that about 68% of the time $\bar{Y}$ will be within ± 60 mg of 500 mg.

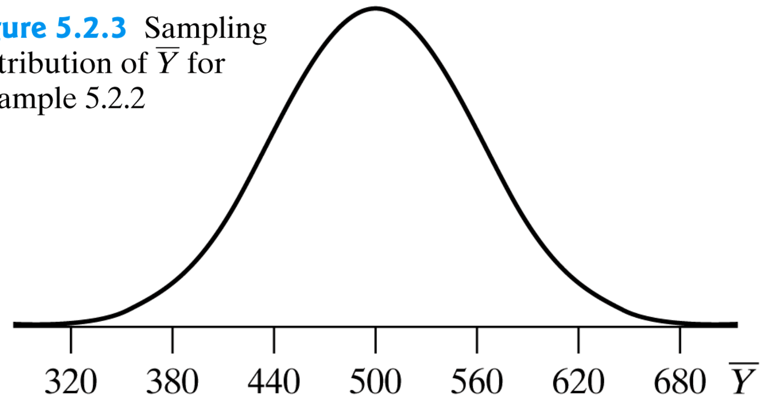


Figure 5.2.3 Sampling distribution of $\bar{Y}$ for Example 5.2.2

5.2 The Sample Mean

Dependence on Sample Size

Consider the possibility of choosing random samples of various sizes from the same population. The sampling distribution of $\overline{\mathrm{Y}}$ will depend on the sample size n in two ways.

1) Its standard deviation is $\boldsymbol{\sigma_{\bar{Y}} = \sigma/\sqrt{n}}$, and this is inversely proportional to √n.
   - **Larger n gives a smaller value of $\boldsymbol{\sigma_{\bar{Y}}}$** , and
   - consequently a smaller expected sampling error if ȳ is used as an estimate of μ.

2) Second, if the population distribution is not normal, then the shape of **the sampling distribution of $\boldsymbol{\bar{Y}}$** depends on n, **being more nearly normal for larger n**.

   * However, if the population distribution is normal, then the sampling distribution of $\overline{\mathrm{Y}}$ is always normal, and only the standard deviation depends on n.

- The more important of the two effects of sample size is the first.

5.2 The Sample Mean

Dependence on Sample Size

Example 5.2.2 Weights of seeds (continued)

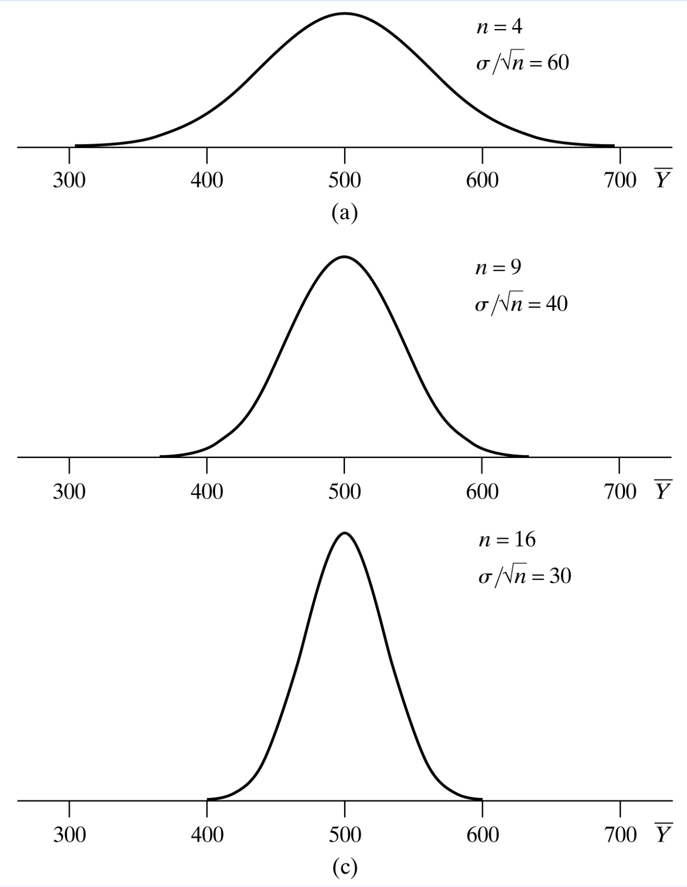
- the population follow a normal distribution with **μ = 500** mg and **σ = 120** mg; n = 4 or 9 or 16.
- What is the standard deviation of sampling distribution of $\overline{Y}$?
- What is the effect of sample size?

5.2 The Sample Mean

Dependence on Sample Size

Example 5.2.2 Weights of seeds (continued)

- the population follow a <u>normal distribution</u> with **μ = 500** mg and **σ = 120** mg; n = 4 or 9 or 16.
- What is the standard deviation of sampling distribution of $\bar{Y}$?
  - According to Theorem 5.2.1, the sampling distribution of $\bar{Y}$ will be a normal distribution.
  - $\boldsymbol{\mu_{\bar{Y}}} = \boldsymbol{\mu}$ = 500 mg
  - $\boldsymbol{\sigma_{\bar{Y}}} = \boldsymbol{\sigma/\sqrt{n}}$ = 120/ √4 = 60 mg
  - $\boldsymbol{\sigma_{\bar{Y}}} = \boldsymbol{\sigma/\sqrt{n}}$ = 120/ √9 = 40 mg
  - $\boldsymbol{\sigma_{\bar{Y}}} = \boldsymbol{\sigma/\sqrt{n}}$ = 120/ √16 = 30 mg
- What is the effect of sample size?

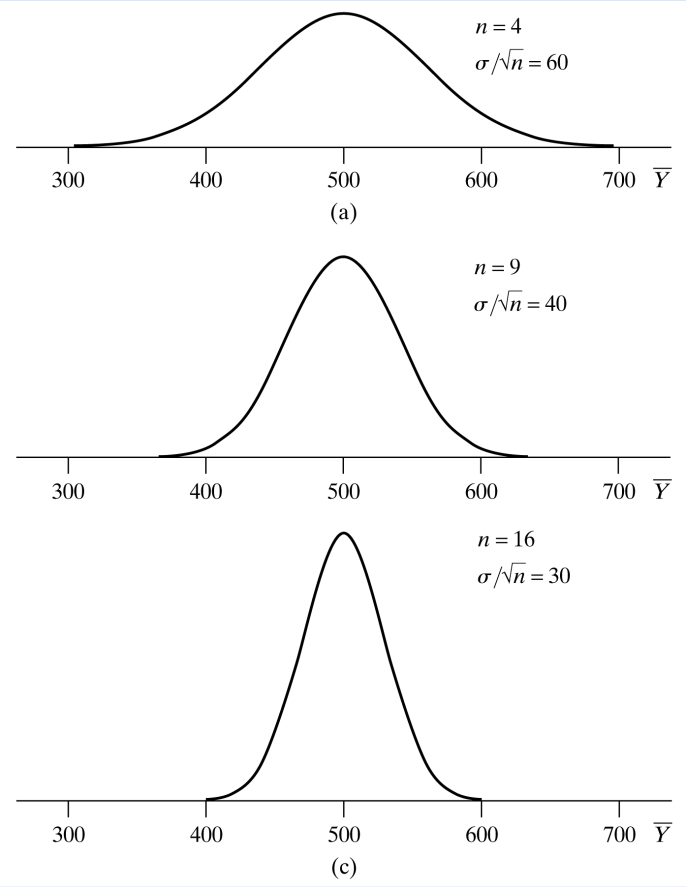


5.2 The Sample Mean

Dependence on Sample Size

Example 5.2.2 Weights of seeds (continued)

- the population follow a normal distribution with **μ = 500** mg and **σ = 120** mg; n = 4 or 9 or 16.
- What is the standard deviation of sampling distribution of $\bar{Y}$?
- What is the effect of sample size?
  - Notice that for larger n the sampling distribution is more concentrated around the population mean μ.
  - As a consequence, the probability that is $\bar{Y}$ close to μ is larger for larger n.
  - In this sense that a larger sample provides more information about the population mean than a smaller sample.



5.2 The Sample Mean

Populations, samples, and sampling distributions

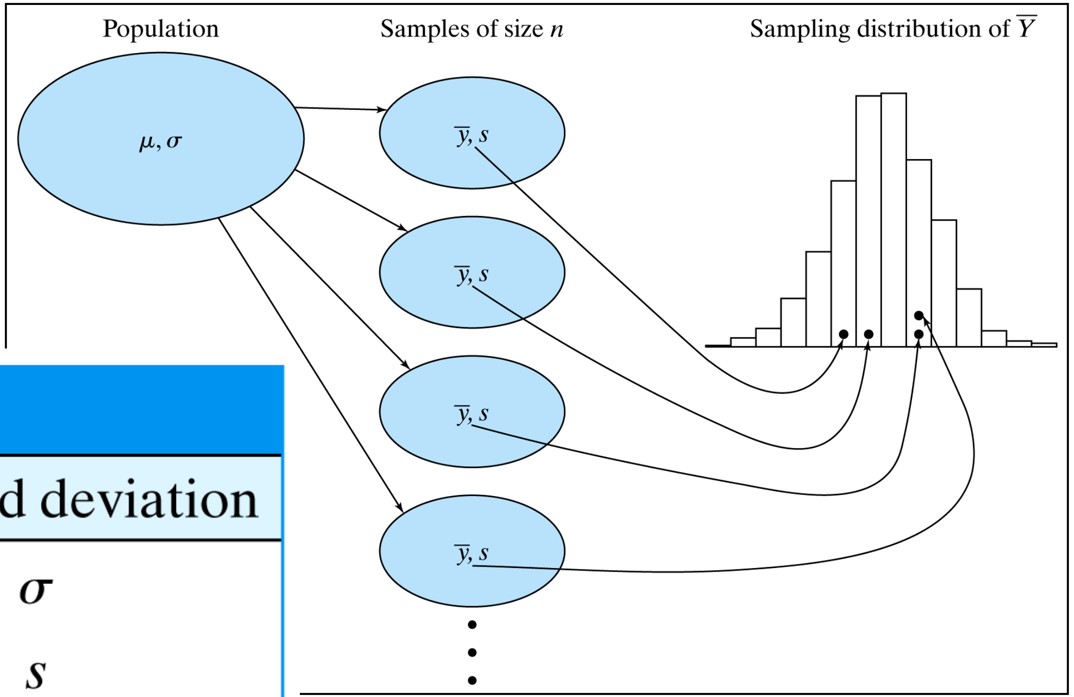


Table 5.2.2

Distribution	Mean	Standard deviation
Y in population	μ	σ
Y in sample	$\bar{y}$	s
$\bar{Y}$ (in meta-study)	$\mu_{\bar{Y}} = \mu$	$\sigma_{\bar{Y}} = \dfrac{\sigma}{\sqrt{n}}$

5.2 The Sample Mean

Other Aspects Of Sampling Variability

- The preceding discussion has focused on sampling variability in the sample mean, $\overline{Y}$.
- Two other important aspects of sampling variability are
  1) sampling variability in the <u>sample standard deviation</u>, s.
  2) sampling variability in the <u>shape</u> of the sample, as represented by the sample histogram.

Example 5.2.2 Weights of seeds

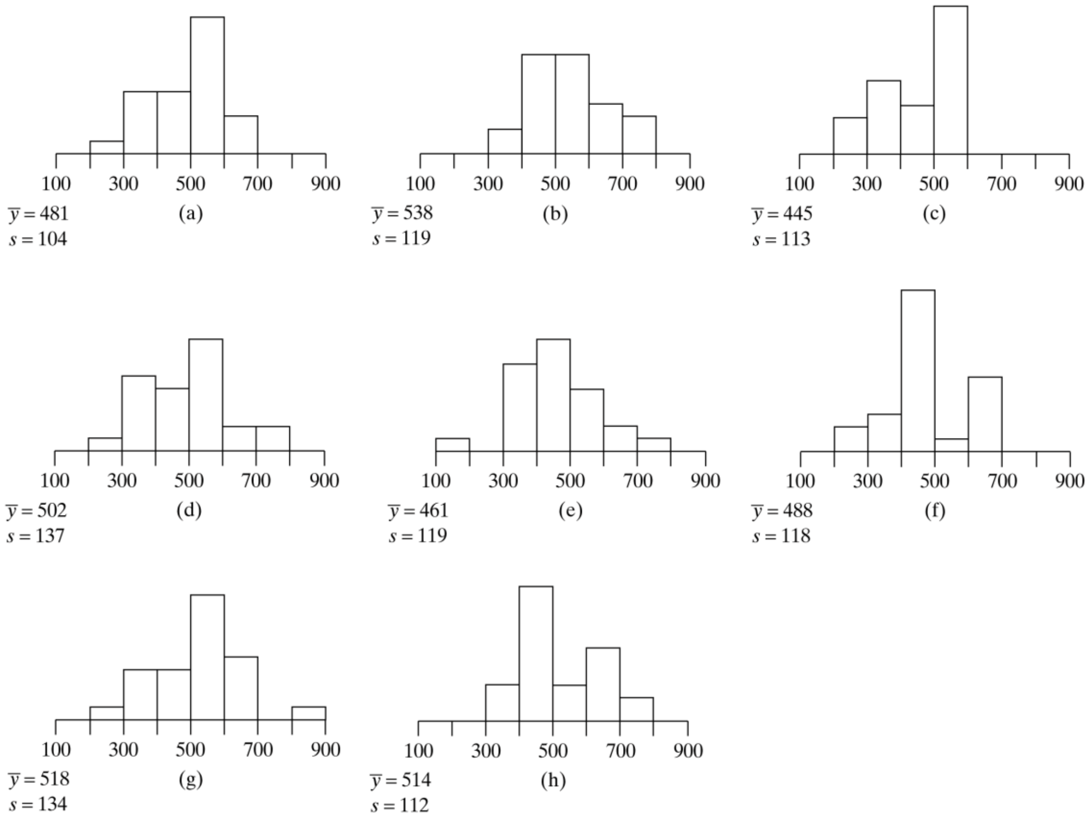


Figure 5.2.7 Eight random samples, each of size $n = 25$, from a normal population with $\mu = 500$ and $\sigma = 120$

Summary

Chapter 5 – Sampling Distribution

- 5.1 Basic Ideas
- 5.2 The Sample Mean

Homework

Chapter 5

- **5.1.1 ; 5.1.5 ;**
- **5.2.4 ; 5.2.14; 5.2.19.**