

This work is licensed under a [Creative Commons Attribution-NonCommercial-ShareAlike License](#). Your use of this material constitutes acceptance of that license and the conditions of use of materials on this site.



Copyright 2007, The Johns Hopkins University and William A. Reinke. All rights reserved. Use of these materials permitted only in accordance with license rights granted. Materials provided "AS IS"; no representations or warranties provided. User assumes all responsibility for use, and all liability related thereto, and must independently review all materials for accuracy and efficacy. May contain materials owned by others. User is responsible for obtaining permissions for use from third parties as needed.

Session 3

Sampling Design Alternatives

William A. Reinke, Ph.D.

Professor

Department of International Health

Johns Hopkins University

School of Hygiene and Public Health

Principles To Be Developed

- **Sample Statistics Differ from but Are Related to Population Parameters**
- **Difference Can Be Reduced by Obtaining Larger Sample of Data**
- **Some Sampling Designs for Obtaining These Data Are**
 - **More Informative**
 - **Less Costly**
 - **More Efficient**

Main Measures of Interest

Population
Parameter

Sample
Statistics

Continuous Variables

- Average: Arithmetic Mean
- Dispersion: Standard Deviation

μ

σ

\bar{X}

S

Discrete Variables

- Relative Frequency: Proportion

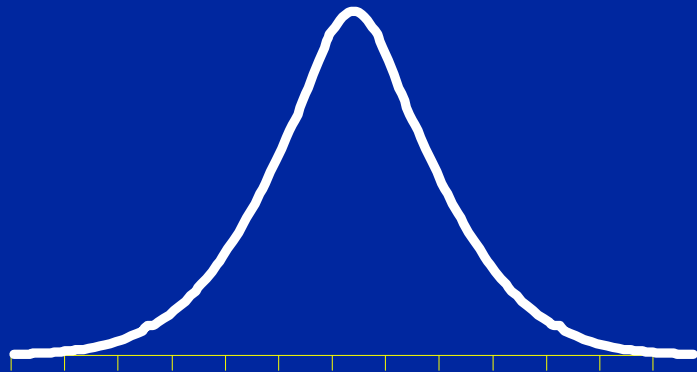
π

P

UNIVERSE



SAMPLE

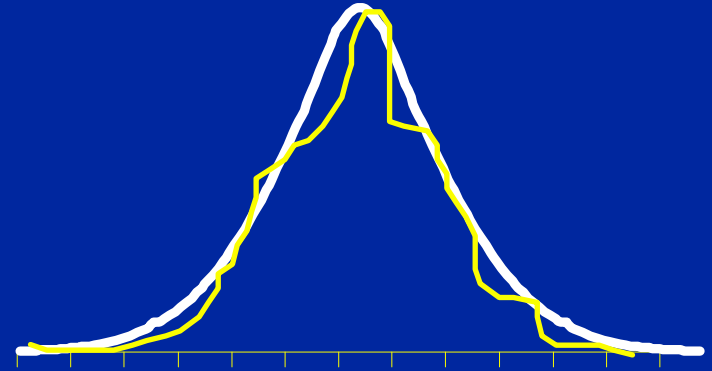


Parameters

μ

σ

π



Statistics

\bar{X}

S

P



Estimates

Hypothetical Sample Results

Three Populations

	A	B	C
	75	73	79
	75	76	79
	75	74	87
	75	78	72
	75	74	66
\bar{X}	75	75	75
μ	75	≈75	?

Precise Estimates are Possible If -

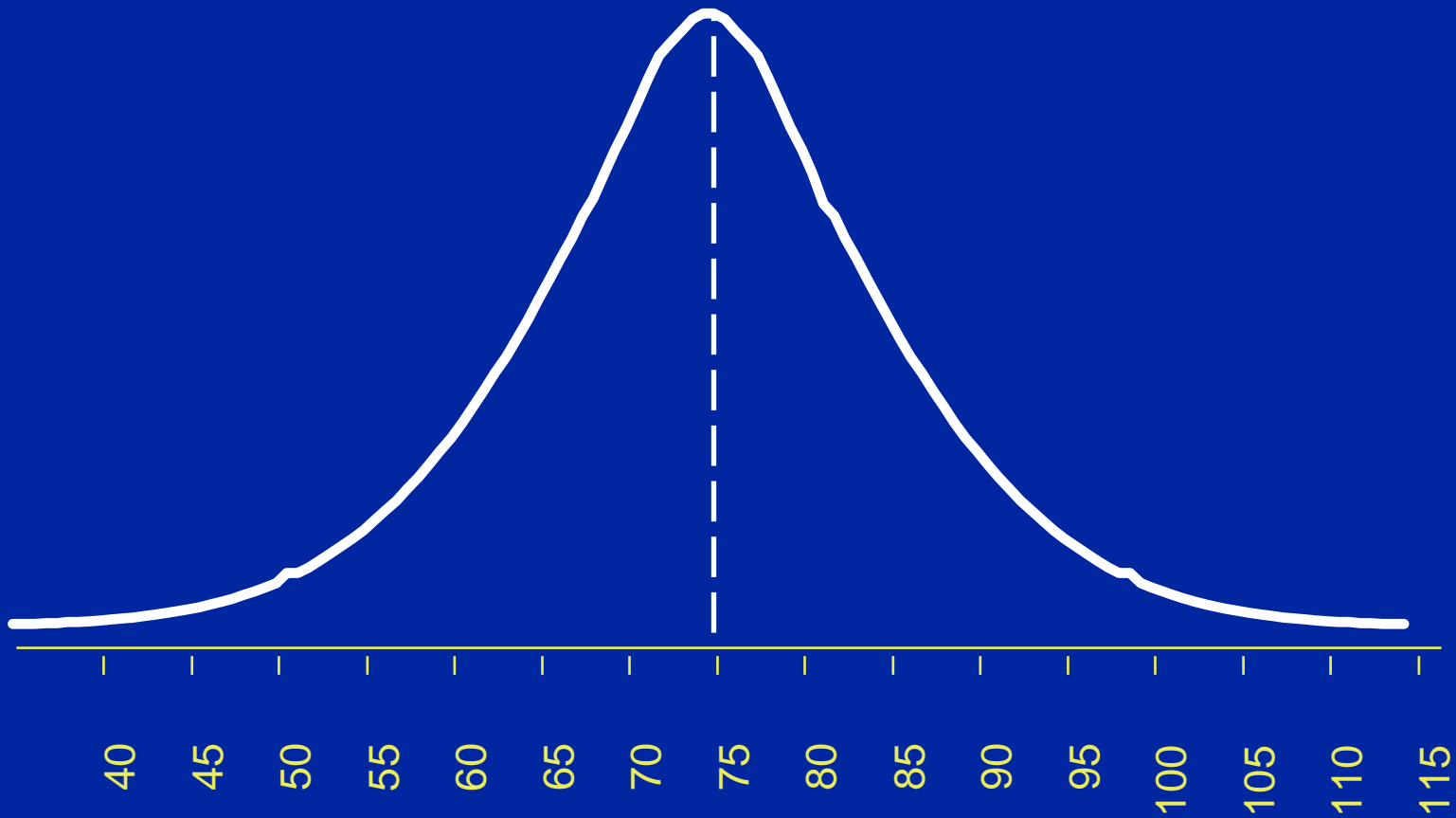
- There is Little Variation Among Sample Results
- The Sample Size is Sufficiently Large

The Mathematical Relationship is -

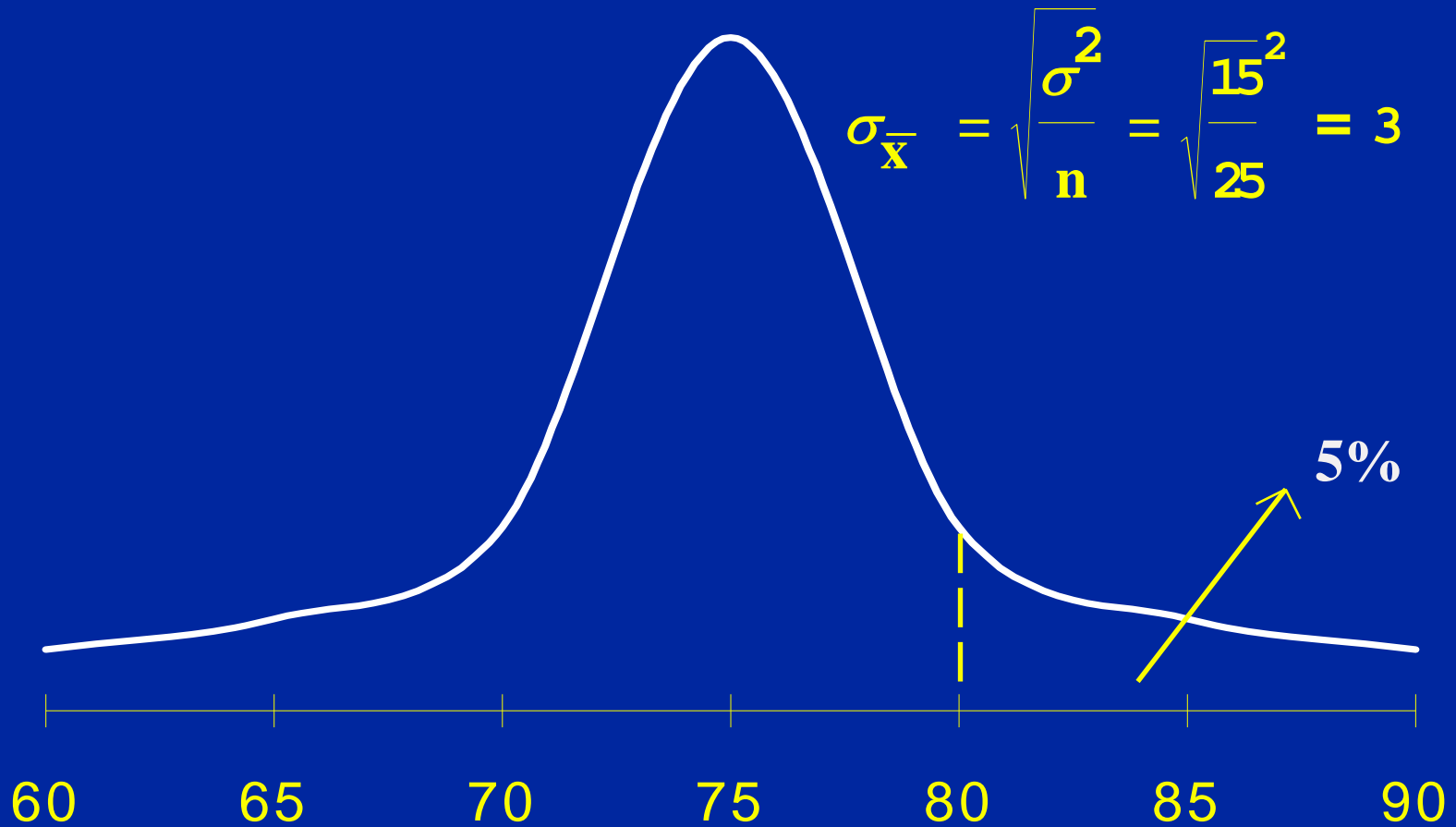
$$\text{Standard Error} = \sqrt{\frac{\text{Variance}}{\text{Sample Size}}}$$

$$= \sqrt{\frac{\sigma^2}{n}} \quad \text{or} \quad \sqrt{\frac{\pi(1-\pi)}{n}}$$

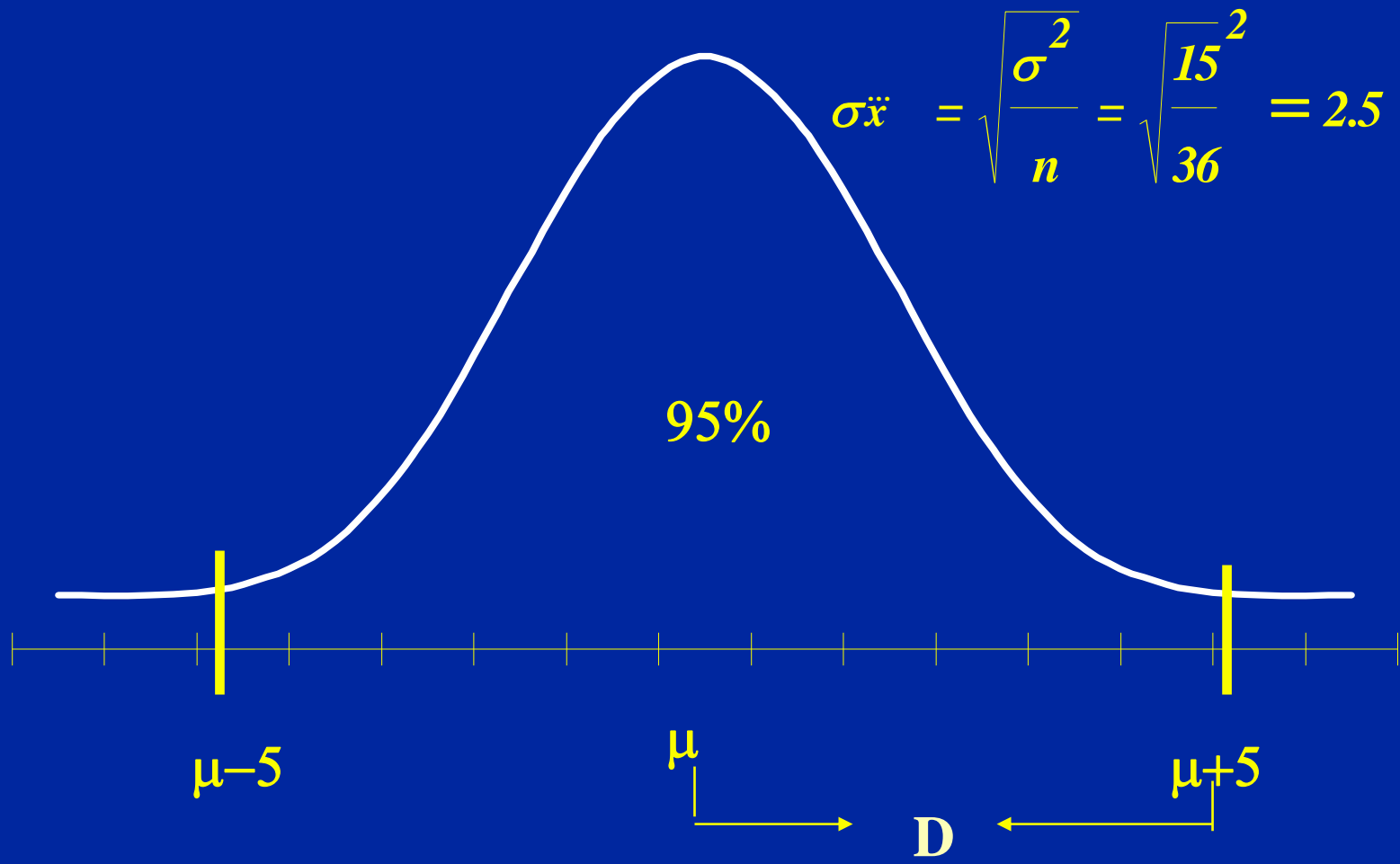
$$\sigma_{\bar{x}} = 15$$



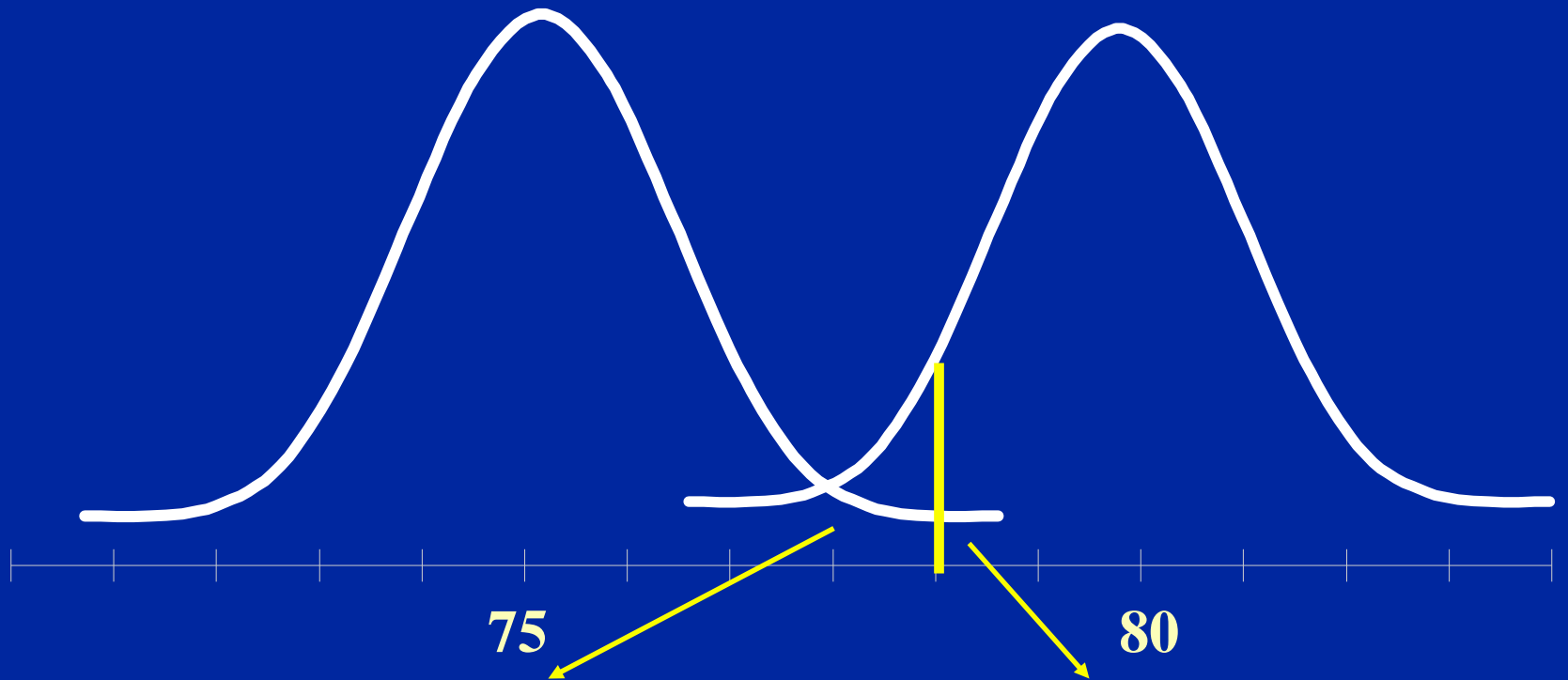
Daily Attendance (X)



Monthly Average of Daily Attendance (X)



Daily Average of 36 Days (X)



Type II Error of Omission
10%

Type I Error of Commission
5%

Determination of Sample Size

Simple Random Sample

Purpose of analysis	Sources of Error	Type of Error	General formula for n	Special case
Estimate universe mean	I	I	$\left[\frac{Z S}{D} \right]^2$	$\frac{4 S^2}{D^2}$
Decide whether Universe Mean Conforms to Defined Standard	I	2	$\left[\frac{(Z_1 + Z_2) S}{D} \right]^2$	$\frac{10.9 S^2}{D^2}$
Estimate Differences between Two Universe Means	2	I	$2 \left[\frac{Z S}{D} \right]^2$	$\frac{8 S^2}{D^2}$
Decide Whether Real (non-zero) Differences Exists between Two Universe Means	2	2	$2 \left[\frac{(Z_1 + Z_2) S}{D} \right]^2$	$\frac{21.8 S^2}{D^2}$

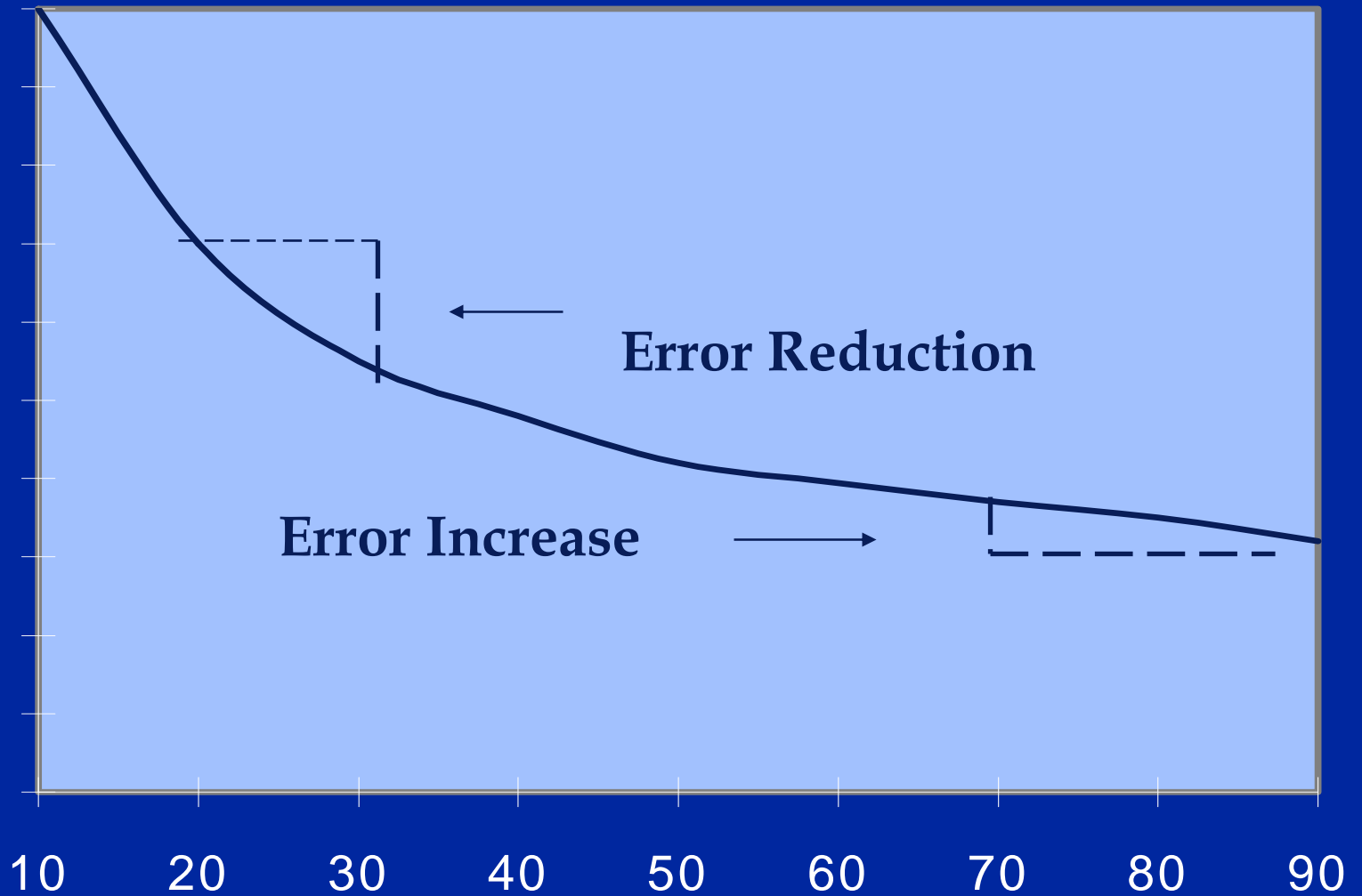
Assumptions:

Z = 2.0 (95% confidence)

Z₁ = 2.0 (5% Risk Type I Error)

Z₂ = 1.3 (10% Risk Type II Error)

**Sampling
Error**



Sample Size

Rules of Stratification

for *Separate Analysis* of Population
Subgroups

- **Select Subgroups as *Homogenous* as Possible**
- **Equalize Subgroup Sample Sizes as Much as Possible**

Population Situation

Subgroup <u>Village</u>	Members per Subgroup <u>(Households)</u>
A	400
B	800
C	200
D	500
E	100
	<hr/> 2,000

Sampling Requirement

- **Sample of 20 Households from Each of 3 Villages**
- **At Start Each of Household Has 60 Chances in 2,000 ($p=.03$) to Be Selected**

Sampling Requirement

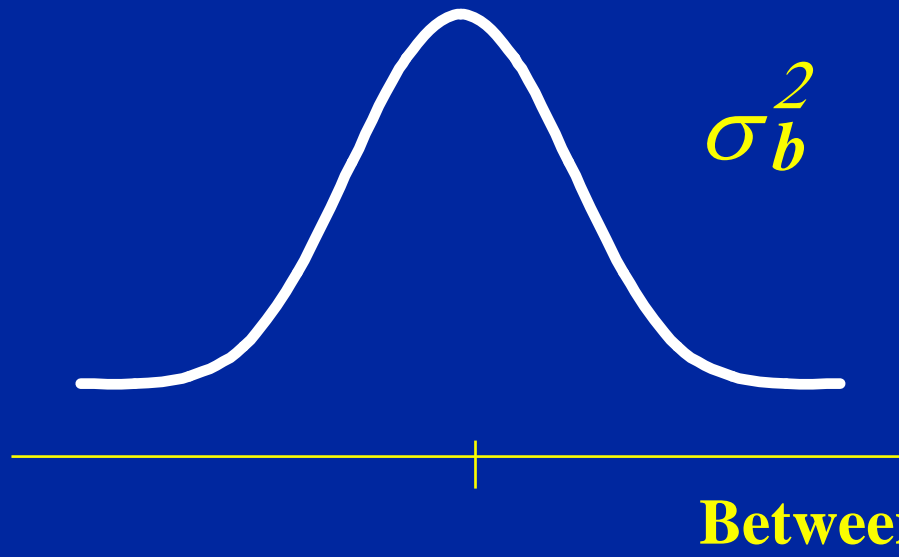
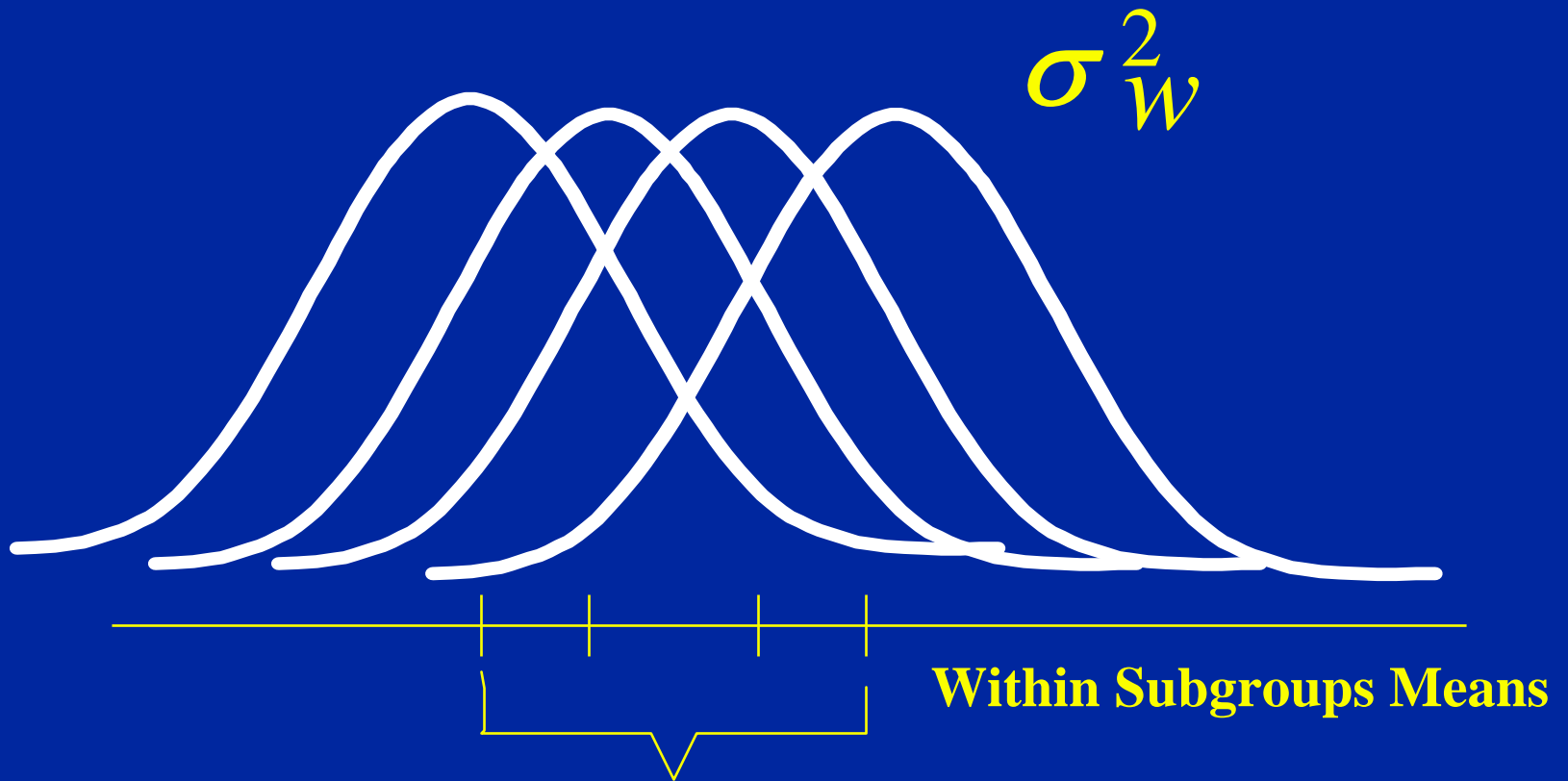
- Sample of 20 Households from Each of 3 Villages
- At Start Each of Household Has 60 Chances in 2,000 ($p=.03$) to Be Selected

Example

Probability that a Specific Household in Village D is Selected:

$$3 \times \frac{500}{2,000} \times \frac{20}{500} = \frac{60}{2,000}$$

Village Chosen	Probability Proportional to Size(PPS)	Probability in Selected Village
----------------	---------------------------------------	---------------------------------



Rules of Multistage Sampling for Combining Subgroup Information to Obtain Aggregate Estimates

- **Select Subgroups as *Heterogeneous* as Possible**
- **Select Subgroups with Probability Proportional to Size (PPS)**
- **Obtain Equal Number of Observations per Subgroups**