Ollscoil Teicneolaíochta an Atlantaigh

Atlantic Technological University

Lean Sigma BIO 08025

Lecture 4

Lecture 4-Define Phase-Surveys and Sampling

Dr. Stephen Daly

Define-Roadmap

- **Identify Problem**-Through the project selection process, the management team identify areas for business improvements

- **Complete Charter**-Project Sponsor leads an effort to document the opportunity in a project charter

- **Validate charter**-The project charter is assigned to a Black Belt who identifies key customers and their requirements, and validates the scope, assumptions and benefits in the charter

Define-Roadmap (contd.)

- **Key Project Charter Validation** steps (Validate or develop the following elements) include:

- **SIPOC** Map-A basic tool for project/process scope definition

- **Gather VOC and VOB information-** Gather the fundamental process performance requirements/needs (Quality, speed, costs) from customers/process owner

- **Develop CCRs and CBRs-** finalize and prioritize Critical Customer (CCRs) and Business (CBRs) requirements

- **Identify Financial benefits-** Understand and describe how the operational improvements to be achieved by the project tie to the financial statements

Define-Roadmap (contd.)

- **Finalize Project Focus-** Modify the charter to incorporate information during project charter validation

- **Project Launch-** A team is selected for the project and the project launched

- **Complete Define gateway Review-** Meet with the project sponsor, business leadership, and other interested parties to ensure alignment of the project focus and management expectations.

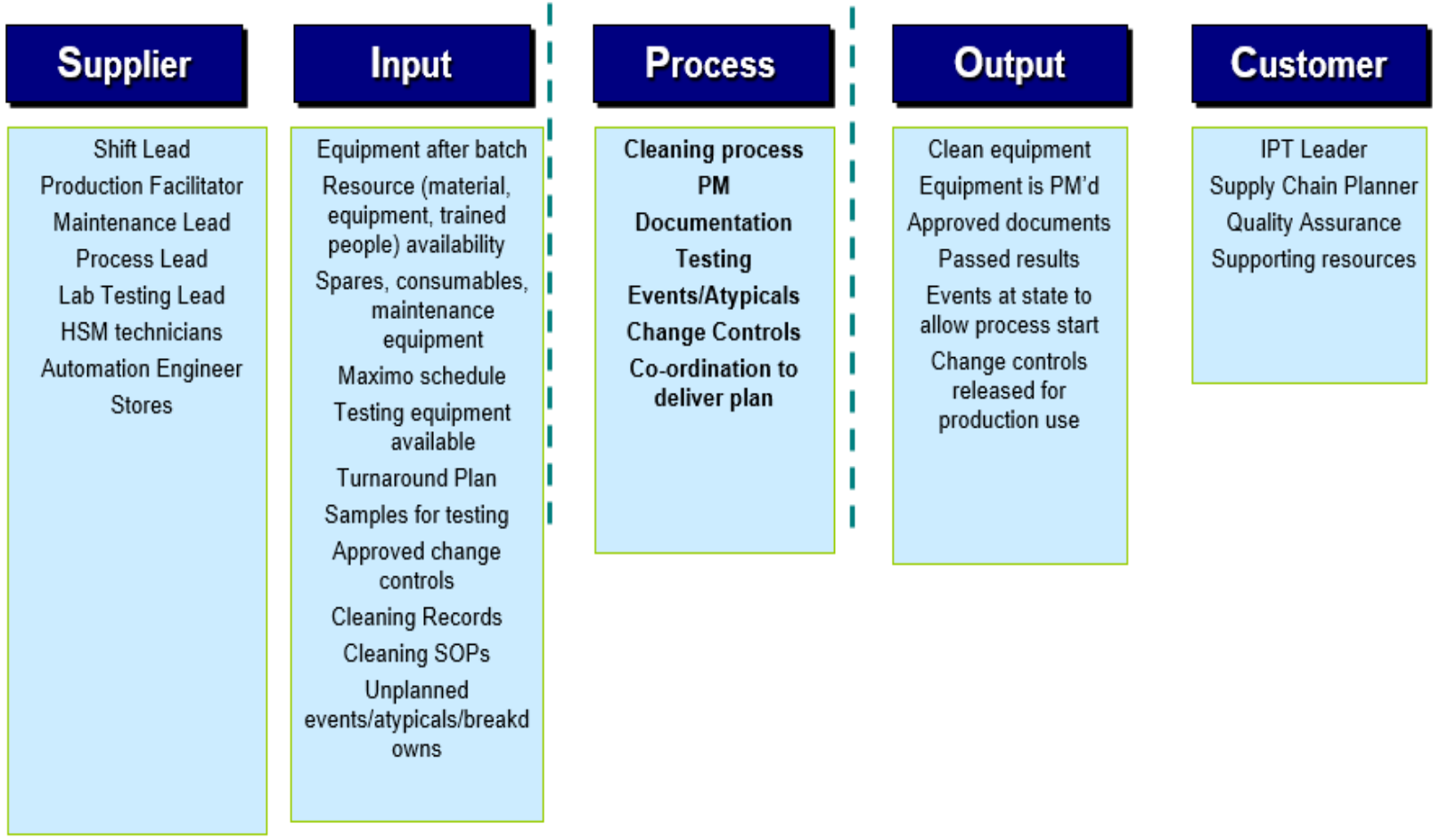
SIPOC

- **S**uppliers
  - Individuals/groups whoever provide whatever is worked on in the process
- **I**nput
  - All inputs to the process i.e. materials, forms, information etc.
- **P**rocess
  - One block showing the steps used to do the work
- **O**utput
  - The product, service, or information being sent to the customer (internal or external)
- **C**ustomers
  - The next step in the process (internal) or the final (external) customers

A high-level SIPOC chart helps visualize the Voice of the Customer and begin to see the relationships between the outputs and the inputs of the process

SIPOC for a pharma cleaning Process

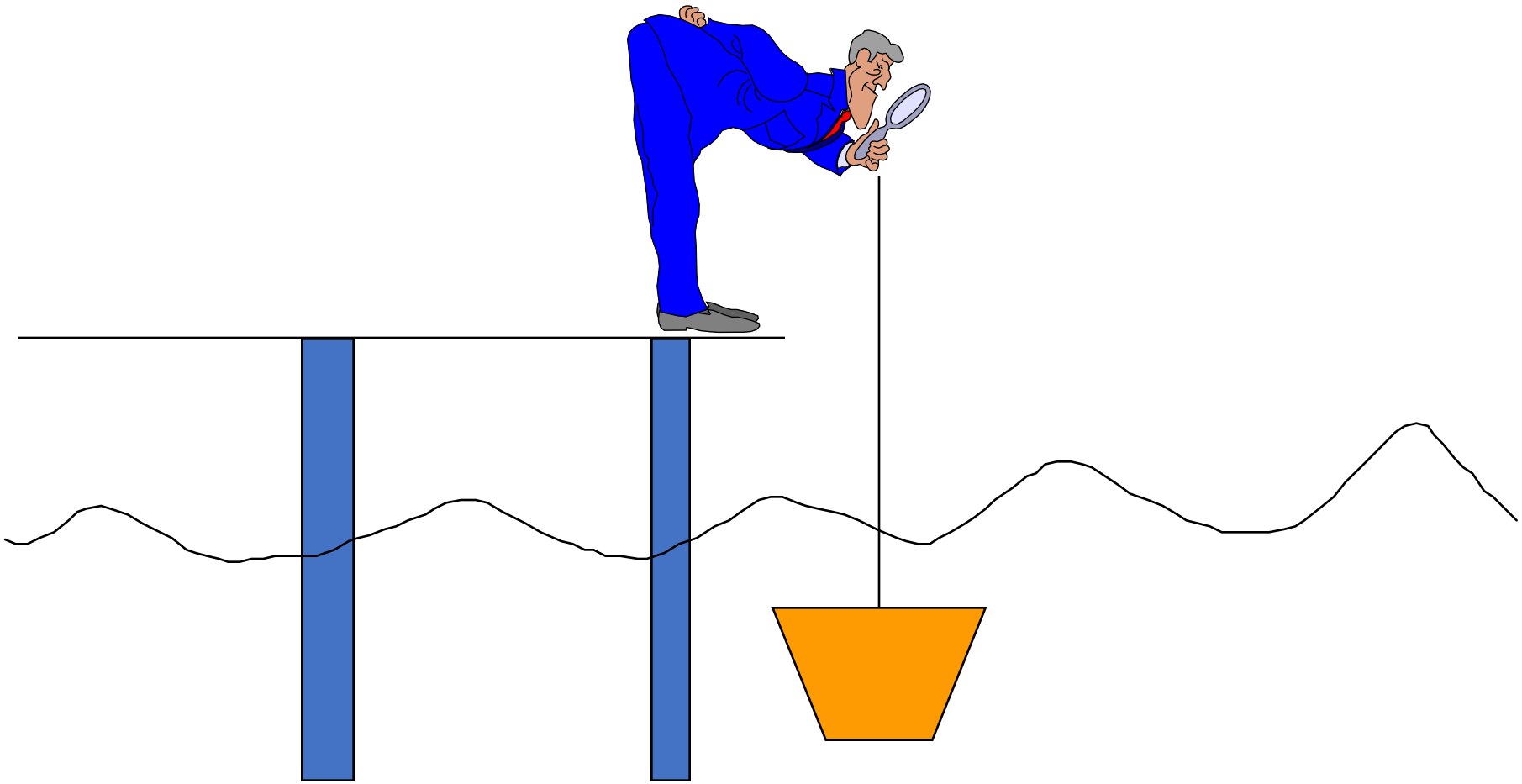
(Project related to reducing variability in cleaning)



Two branches of the science of statistics

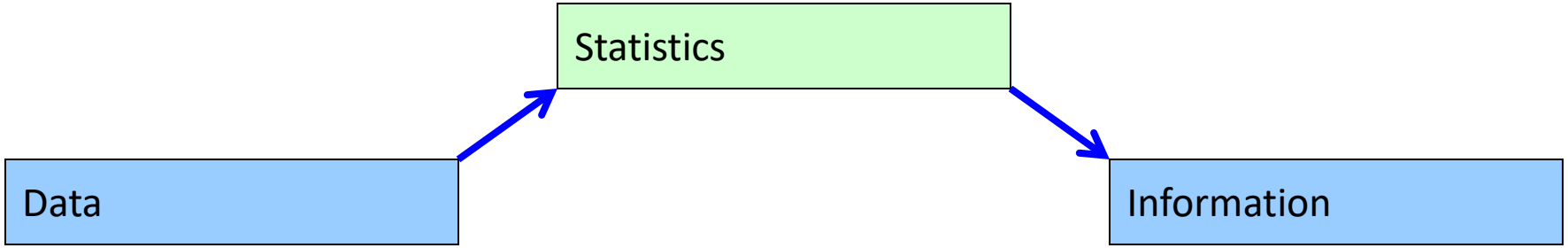
- **Descriptive Statistics**
  - Concerned with describing or characterizing the obtained sample data
  - Use of summary measures—typically measures of central tendency and spread
- **Inferential Statistics**
  - Involves using obtained sample statistics to estimate the corresponding population parameters
  - Most common inference is using a sample mean to estimate a population mean (surveys, opinion polls)

The Concept of Statistical Inference



What is Statistics? Where does this Data come from?

- "Statistics is a way to get information from data"



Data: Facts, especially numerical facts, collected together for reference or information.

Information: Knowledge communicated concerning some particular fact.

Statistics is a ***tool*** for creating ***new understanding*** from a set of numbers.

Population & Samples

- **Population**
  - The complete set of individuals or objects that the investigator is interested in studying
- **Sample**
  - A subset of the population that is actually being studied
- **Variable**
  - A characteristic of an individual or object that can have different values (as opposed to a constant)
- **Independent variable**
  - The variable that is systematically manipulated or measured by the investigator to determine its impact on the outcome.

Population Examples

- Unemployment - Status of ALL employable people (employed, unemployed) in Ireland

- Leaving Cert – Maths grade of EVERY person that took the Honours Maths paper in 2024

- Responses of ALL currently enrolled underage college students in ATU Sligo as to whether they have used public transport in the last 24 hours

Sample Examples

- Unemployment - Status of the 1000 employable people interviewed.
- Leaving Cert Grades - Maths scores of 20 people that took the Honours Leaving Cert Maths paper in 2024.
- Responses of 538 currently enrolled underage college students as to whether they have used public transport in the last 24 hours

Parameters & Statistics

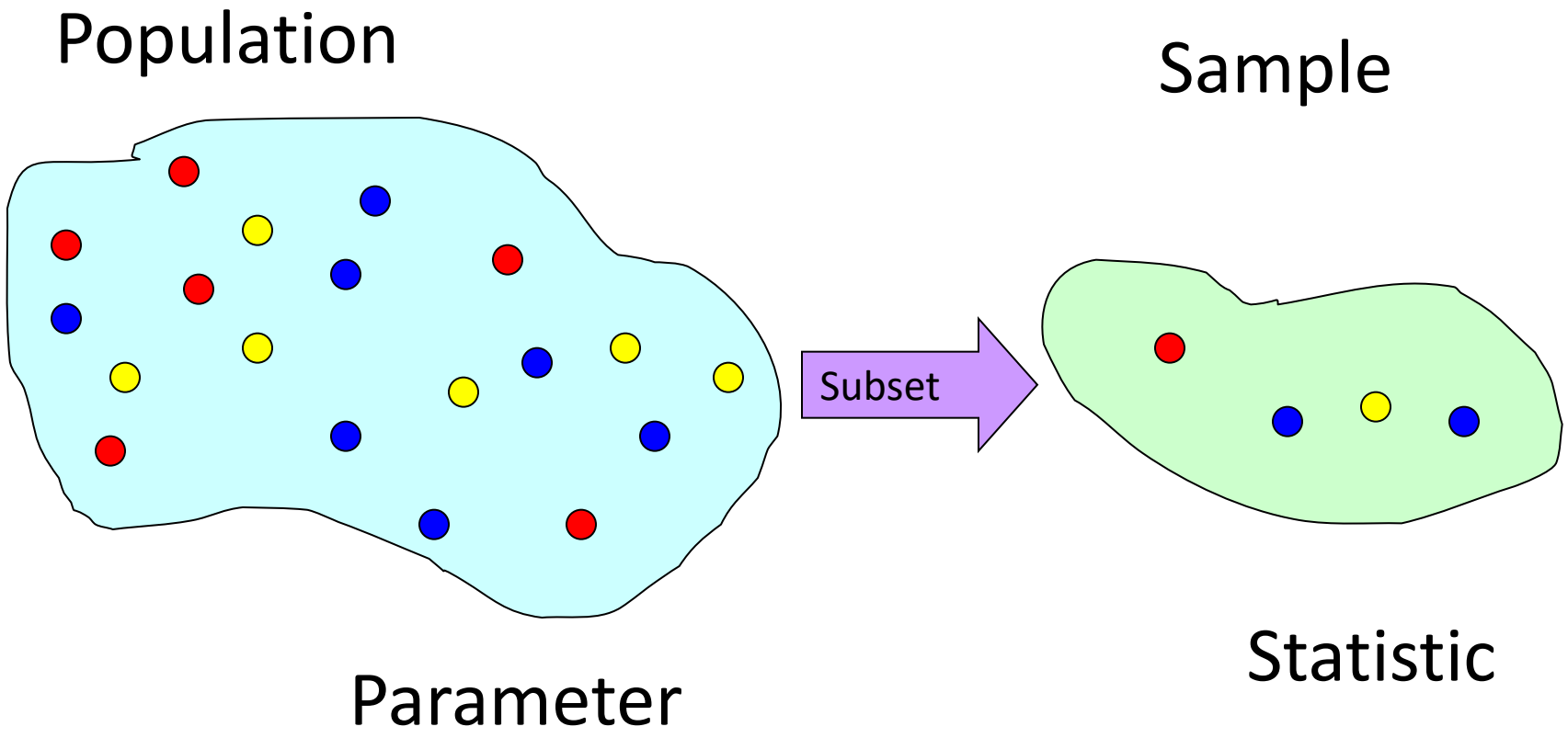
Parameter

- — A descriptive measure of a ***population.***
- *- the true percent of Irish Voters who will vote for a particular candidate – Bart Simpson*

Statistic

- — A descriptive measure of a ***sample.***
- *- Of the 1000 exit voters polled, 550 indicated that they voted for Bart Simpson or 550/1000 = 0.55 or 55%*

Populations and Samples



- **Populations have Parameters,**
- **Samples have Statistics.**

Parameters & Statistics

Variable

1 2 3 (4) 5

Statistic

Average = 3.75

Parameter

Average = 3.72

Quiz - Parameters

A plant manager wants to know if, on average, assembly time for all carburetors is longer than the target time. What is the parameter of interest in the study?

A. Mean assembly time of the entire population

B. Mean assembly time of the sample

C. Target carburetor assembly time

D. All assembly times in the sample

A: A Parameter is a numerical value (such as the mean) that references an entire population.

In this case the mean assembly time of all carburetors. By contrast, similar values collected from representative samples are called statistics.

Quiz – Parameters & Statistics

A restaurant manager wants to determine the average satisfaction rating of his customers. He surveys a random sample of diners and asks them to rate their experience on a scale of 1 to 5, 5 being extremely satisfied. What is the statistic that estimates the manager's parameter of interest?

A. The proportion of satisfied diners among those surveyed

B. The proportion of all satisfied diners

C. The mean level of satisfaction among surveyed diners

D. The mean level of satisfaction among all diners

C: The mean level of satisfaction among surveyed diners. A statistic relies on a sample – in this case, those diners who were surveyed – not the entire population.
Note: He wants to determine the 'AVERAGE' satisfaction rating of his customers.

Statistical Inference

Drawing Conclusions (Inferences) about a Population Based on an examination of a Sample taken from the population

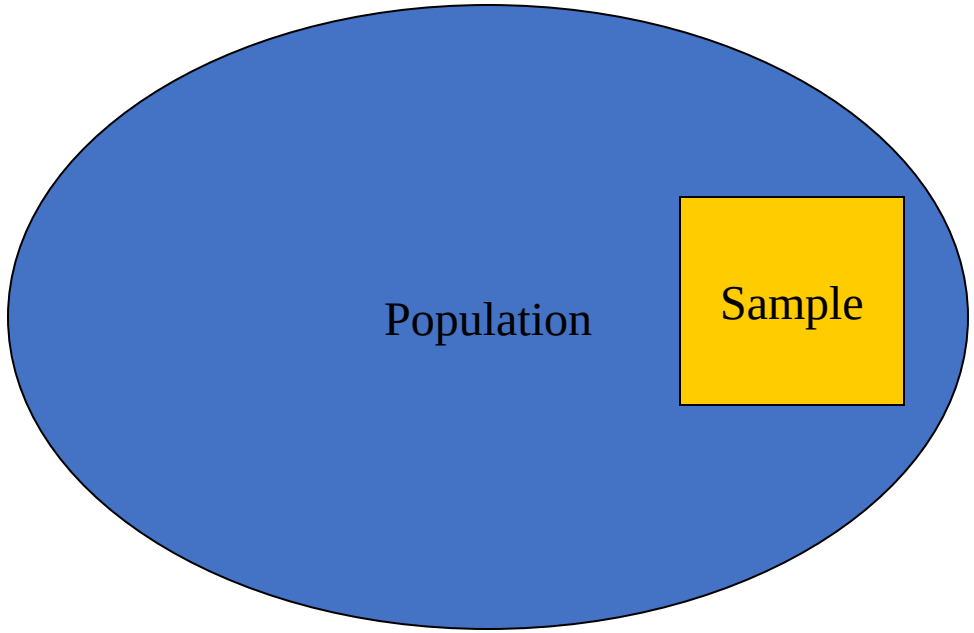
Why use Inferential Statistics?

In most studies, it is difficult to obtain information from the entire population. We rely on samples to make estimates or inferences related to the population.

Statistical Inference Examples

- TV Viewer Ratings
- Market Research
- Financial Auditing
- Opinion Polls
- Incoming Inspection
- Acceptance Sampling

Population vs. Sample



Samples

Again Sample Defined:

A Subset of a population.

A Representative Sample

- **Has the characteristics of the population**

- Census - A Sample that Contains all Items in the Population

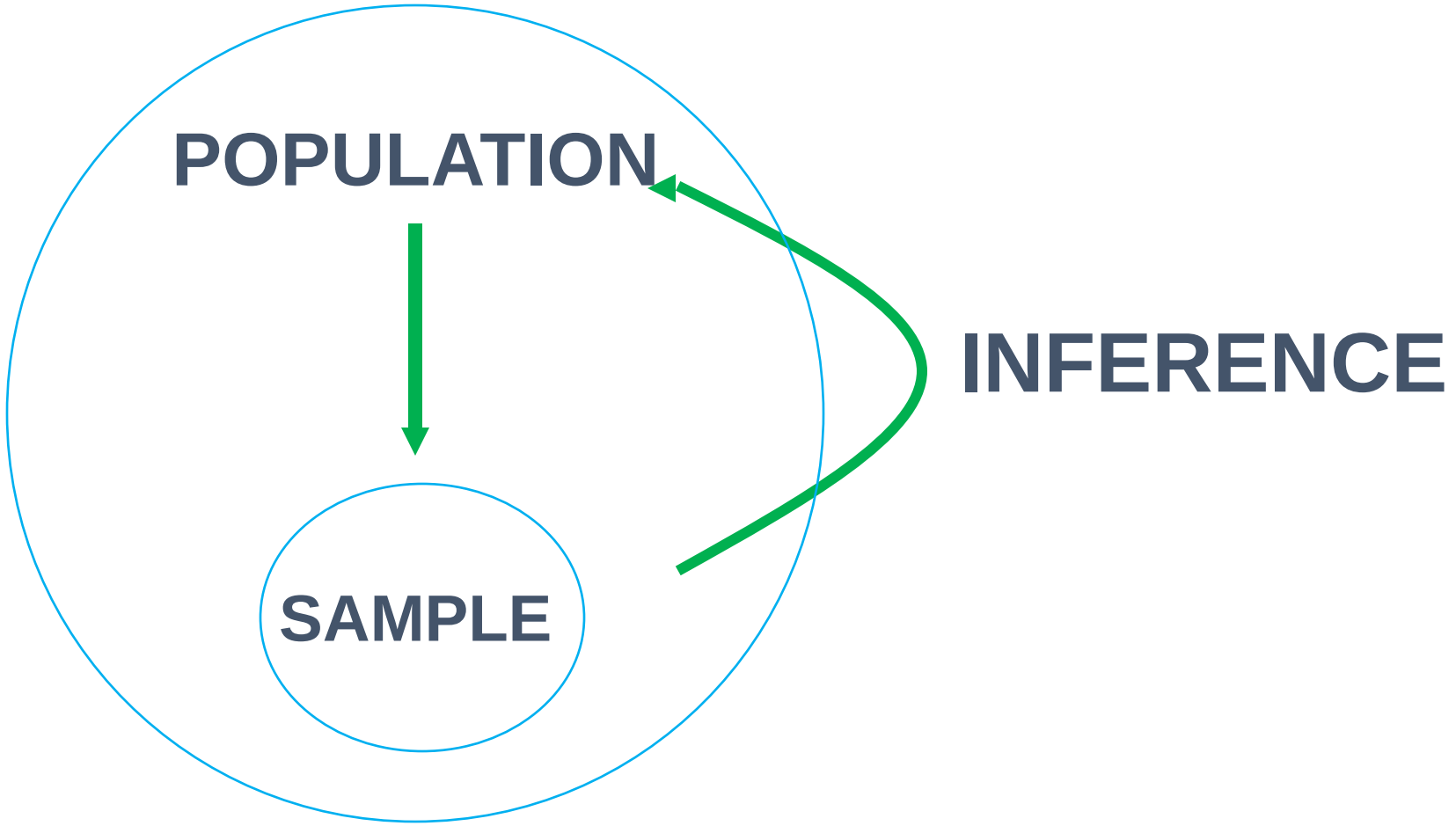
POPULATIONS AND SAMPLES

- Inferential method is based on inferring from a sample to a population
- Sample—a **representative** subset of the population
- Population—the entire set of participants of interest
- Generalizability—the ability to infer population characteristics based on the sample

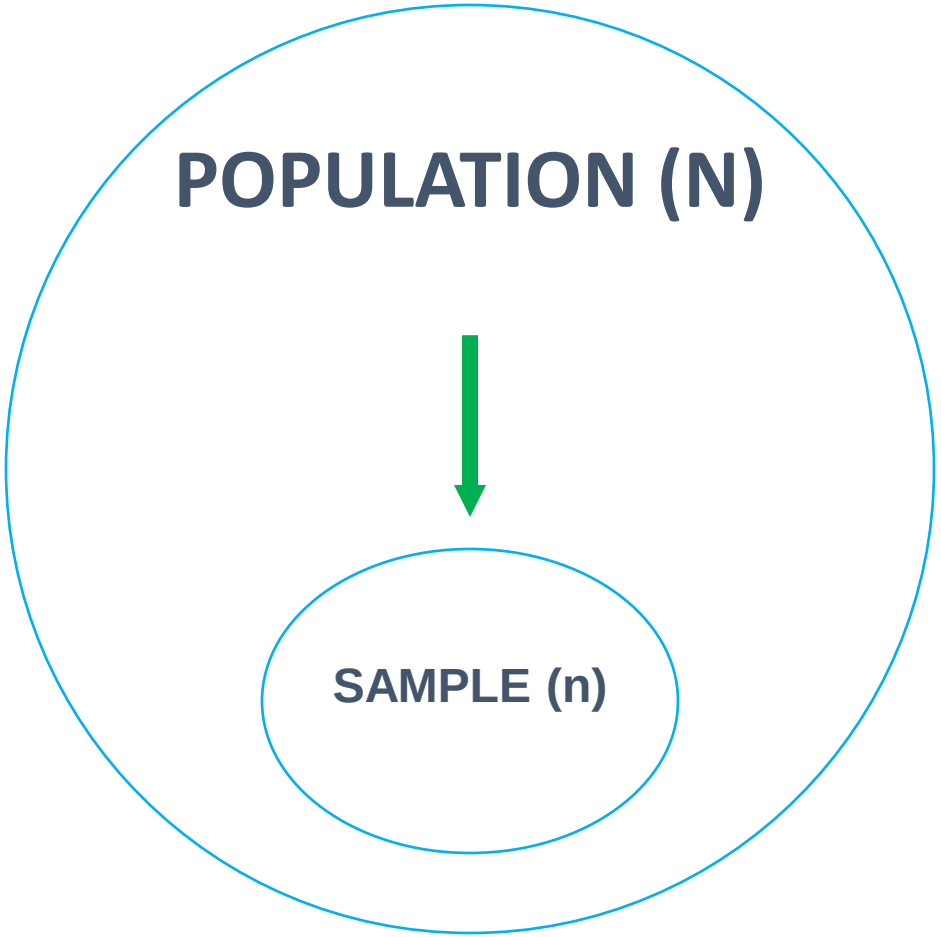
- You may be familiar with the long accepted practice in industry of product Acceptance sampling. i.e. the process of evaluating a portion of the product in a lot for the purpose of accepting or rejecting the entire lot.

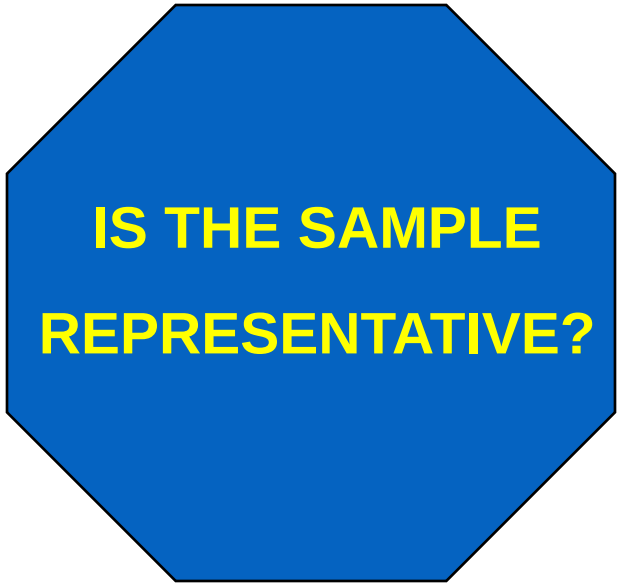
- The main advantage of sampling is cost reduction. The reality however is that it may be impossible to test the entire population. Therefore, it is important that the sample selected is an accurate representation of the population.

The Sampling Process…



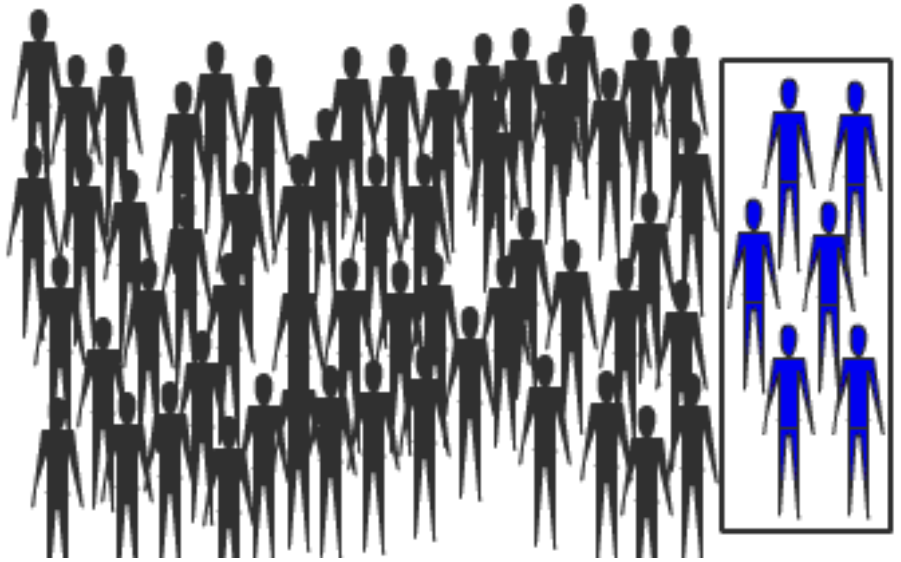
Regarding the sample…



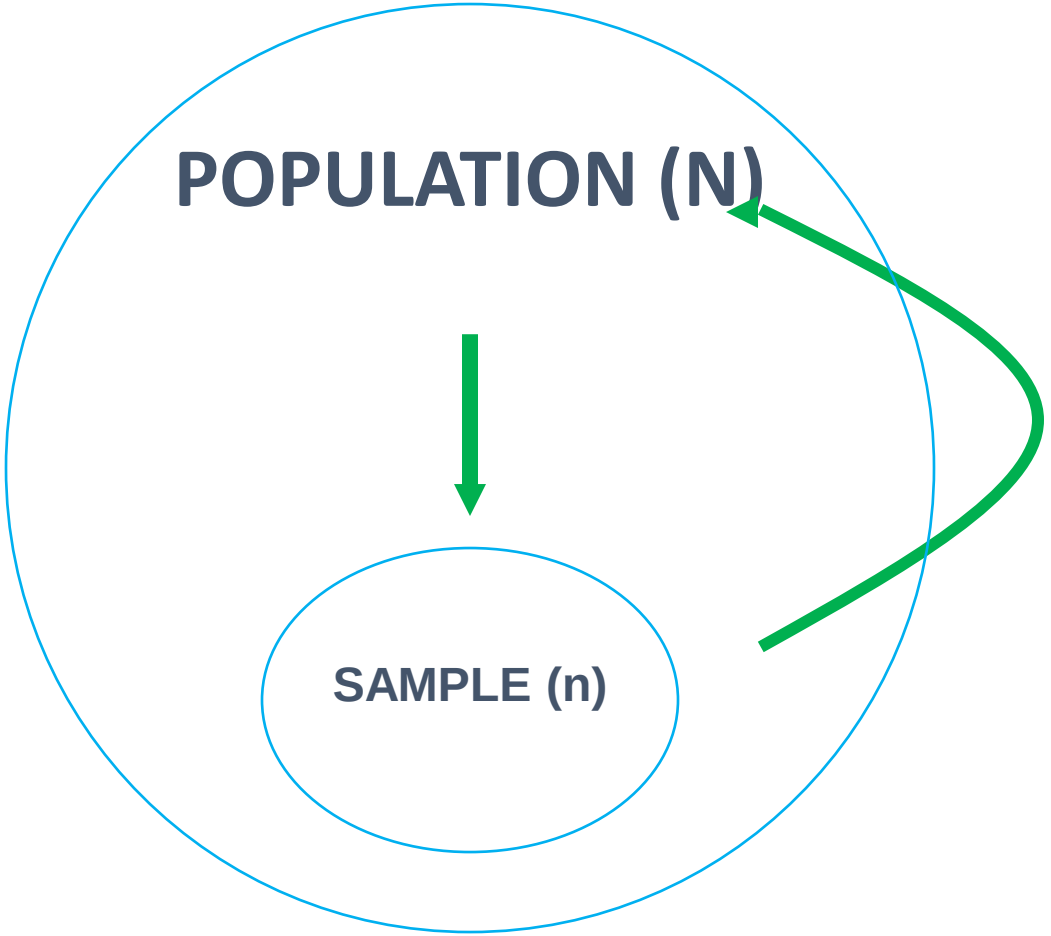


Representative Samples

- The process of selecting a number of individuals for a study in such a way that the individuals represent the larger group from which they were selected



Regarding the inference…



INFERENCE

IS THE INFERENCE GENERALISABLE?

Quiz - Statistical Inference

A bottling plant fills containers to hold 16 fluid ounces of cola. A plant manager wants to study the process's ability to fill bottles with the target volume of cola. Why should the manager measure a sample of filled bottles rather than the whole population?

A. The population of filled bottles may contain bottles with unusually high or low volumes of cola.

B. Volume measurements from a sample may give more accurate results.

C. It is too costly and time consuming to examine the entire population of filled bottles.

D. Volume data from a sample of filled bottles are easier to interpret than volume data from the entire population.

Steps in sampling…

1. Define population (N) to be sampled

2. Determine sample size (n)

3. Control for bias and error

4. Select sample

1. Define population to be sampled…

- Identify the group of interest and its characteristics to which the findings of the study will be generalised

…called the "target" population (the ideal selection)

…often times the "accessible" or "available" population must be used (the realistic selection)

2. Determine the sample size…

- The size of the sample influences both the representativeness of the sample and the statistical analysis of the data

…larger samples are more likely to detect a difference between different groups

…smaller samples are more likely not to be representative

ESTIMATING SAMPLE SIZE

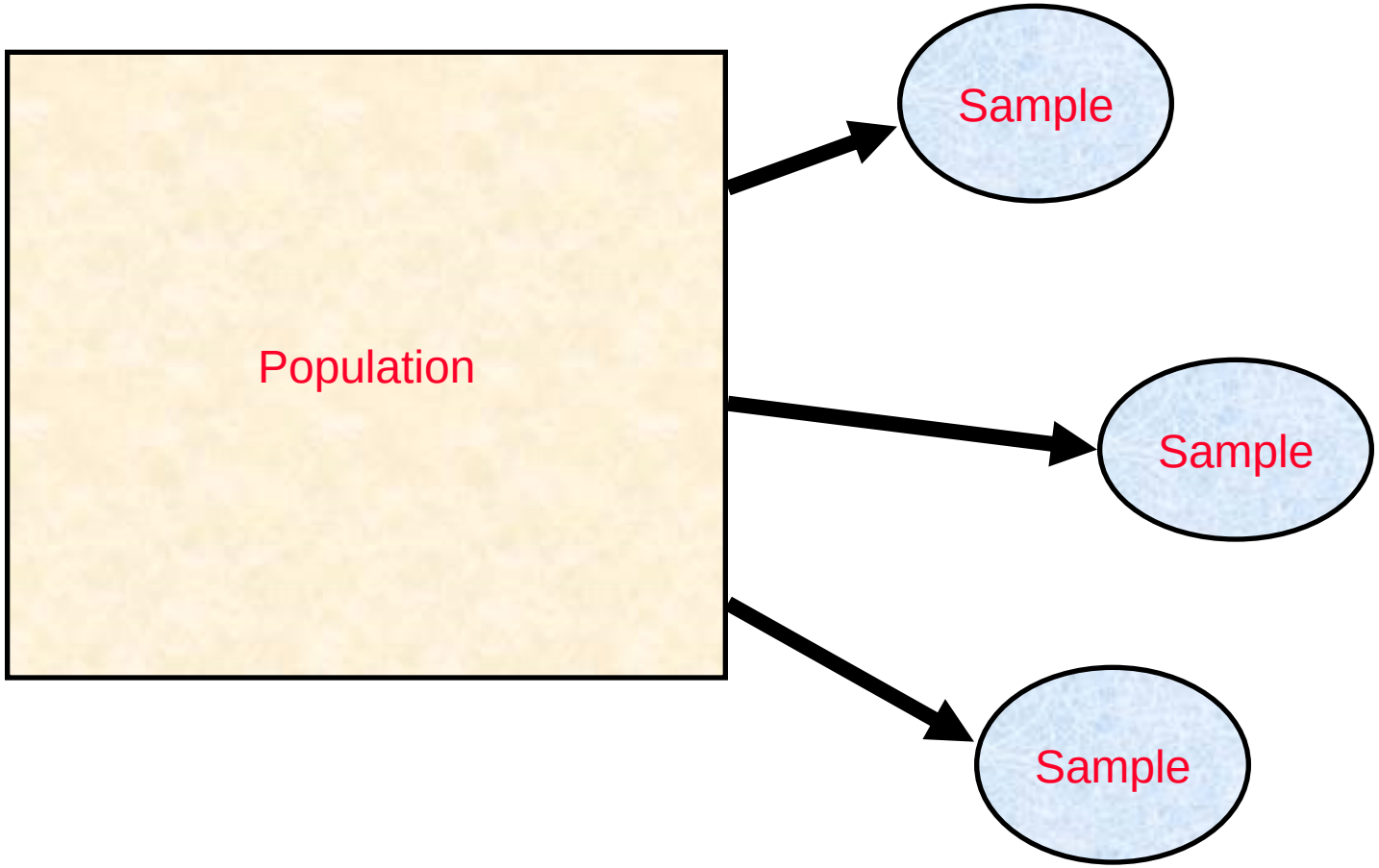
- Generally, larger samples are needed when
  - Variability within each group is great
  - Differences between groups are smaller

- Because
  - As a group becomes more diverse, more data points are needed to represent the group
  - As the difference between groups becomes smaller, more participants are needed to reach “critical mass” to detect the difference

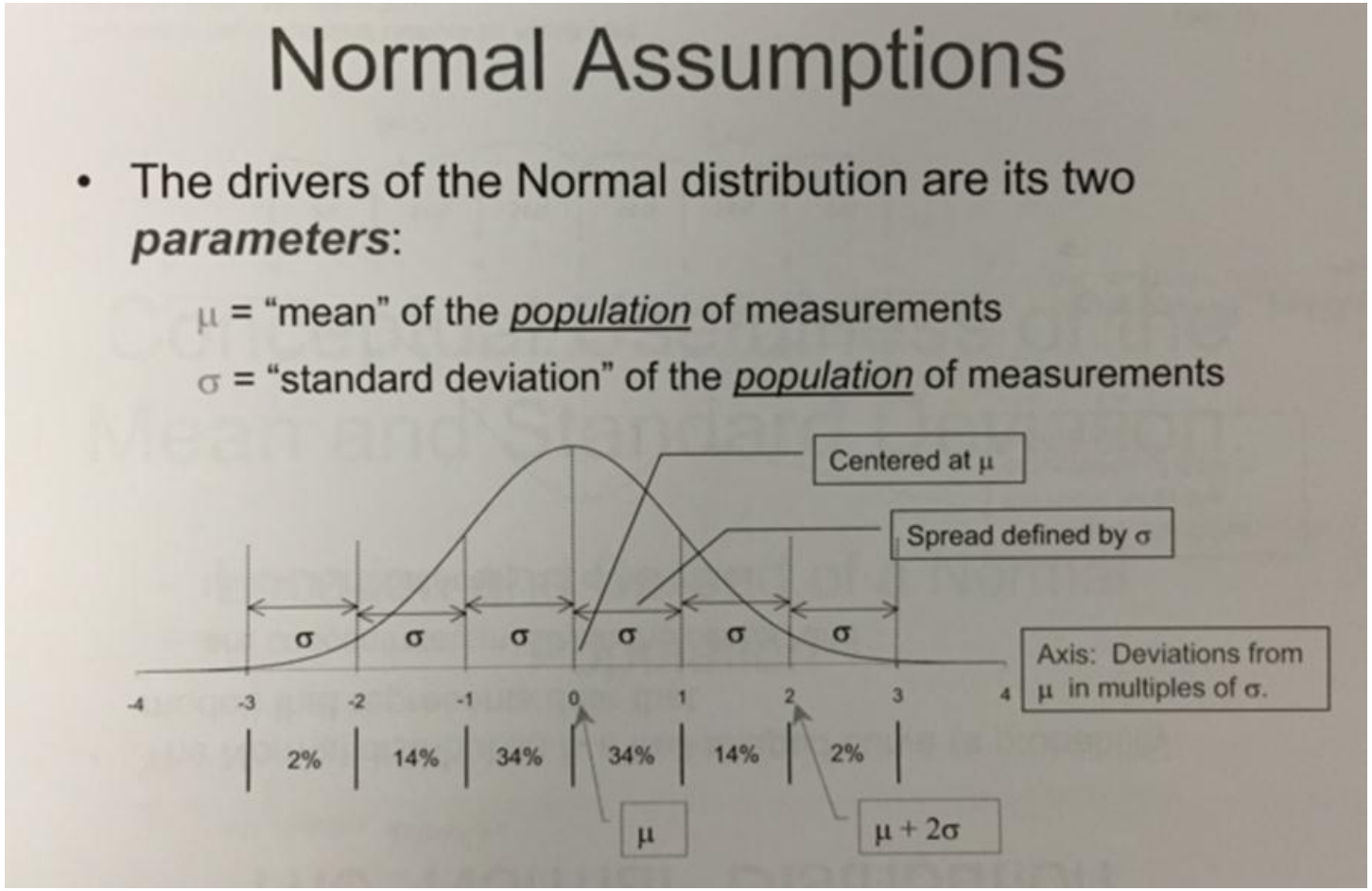
SAMPLES, SAMPLE SIZE, AND SAMPLING ERROR

- Sampling error = difference between sample and population characteristics
- Reducing sampling error is the goal of any sampling technique
- As sample size increases, sampling error decreases

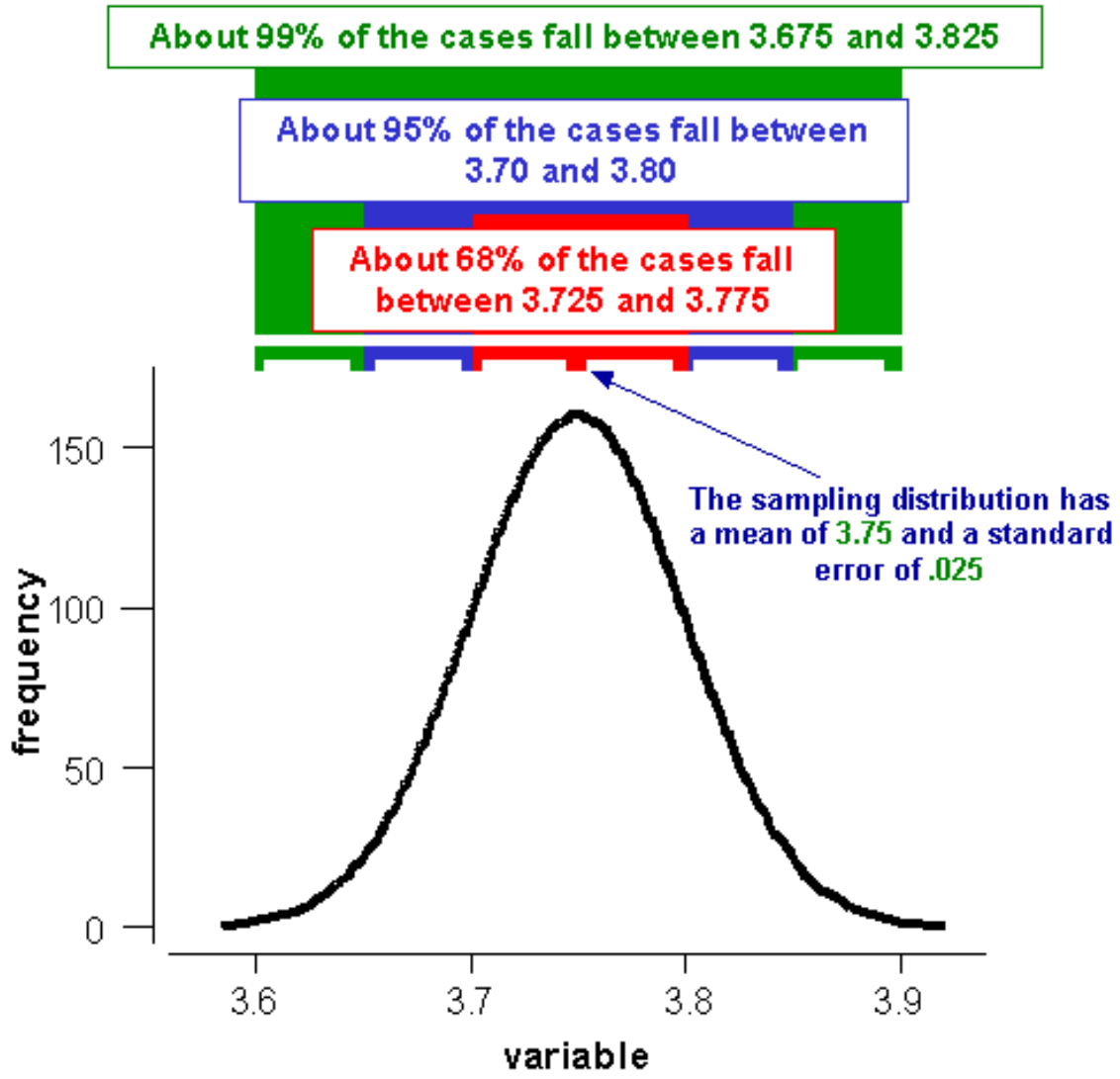
Populations and Samples



Populations and Samples



Sampling Error

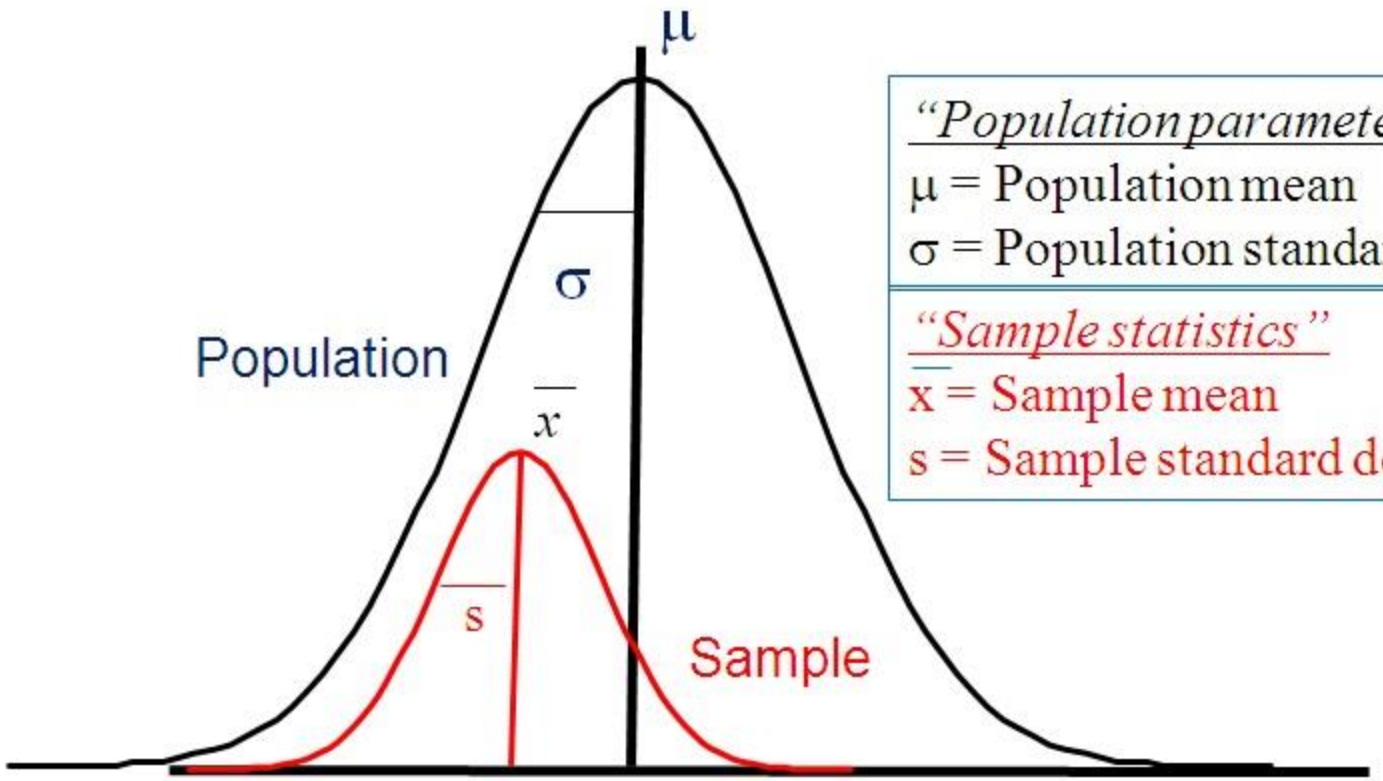


Let's assume we did a study and drew a single sample from the population.

Let's further assume that the average for the sample was 3.75 and the standard deviation was 0.25 shown here.

If we are dealing with raw data and we know the mean and standard deviation of a sample, we can *predict* the intervals within which 68, 95 and 99% of our cases would be expected to fall. We call these intervals the -- guess what -- 68, 95 and 99% confidence intervals

Sampling Error



"Population parameters"

μ = Population mean

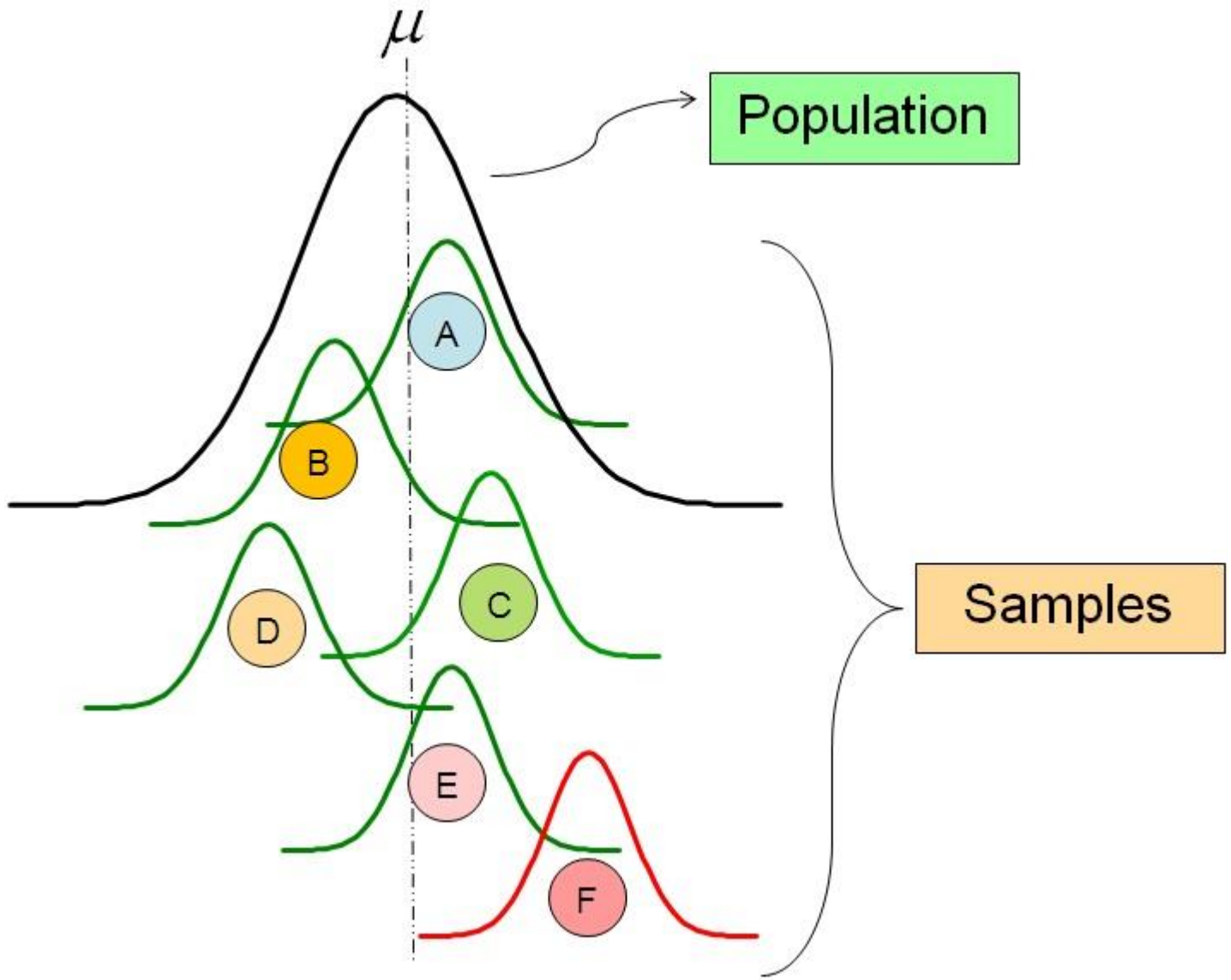
σ = Population standard deviation

"Sample statistics"

$\bar{x}$ = Sample mean

s = Sample standard deviation

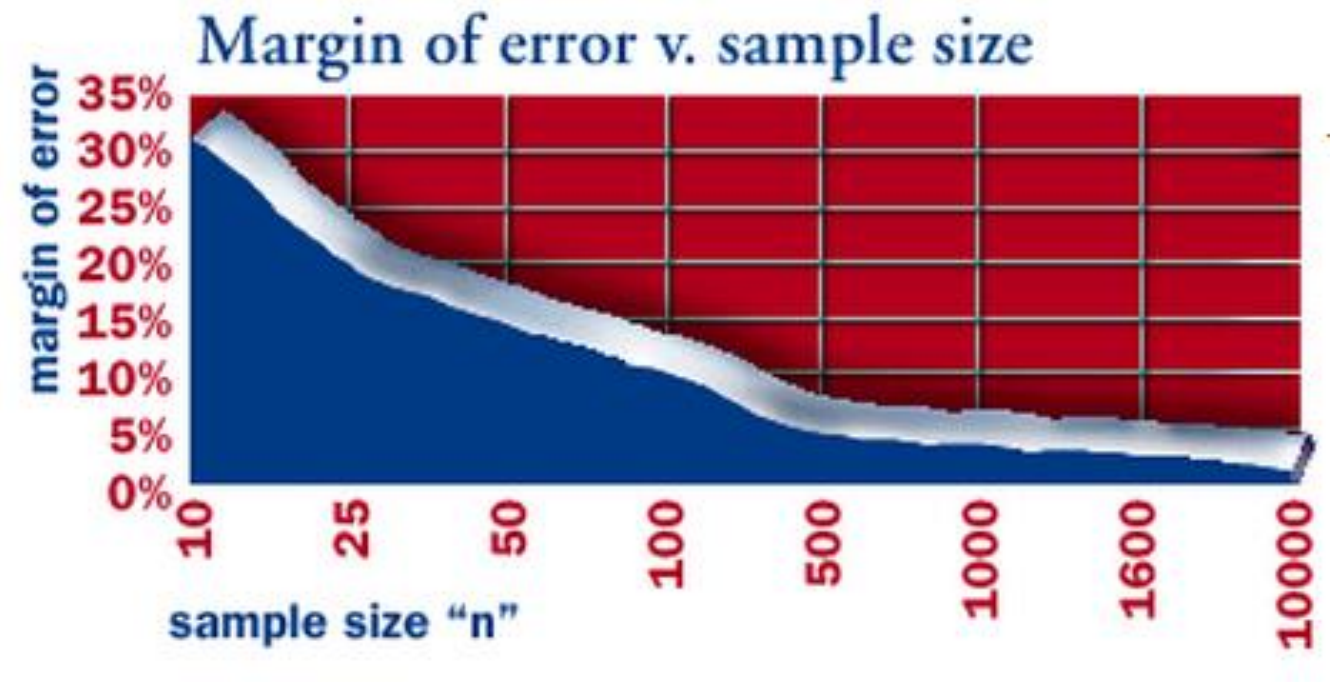
Sampling Error



HOW BIG IS BIG?

- The goal is to select a representative sample—
  - Larger samples are usually more representative
  - But larger samples are also more expensive
  - And larger samples ignore the power of scientific inference

Sampling Error



Going from 100 to 1,000 will decrease your margin of error from 10% to 3%.
But going from 1,000 to 10,000 cases will only reduce the margin to 1%

Rules of thumb for determining the sample size...

1. The larger the population size, the smaller the percentage of the population required to get a representative sample

2. For smaller samples (N ‹ 100), there is little point in sampling. Survey the entire population.

Rules of thumb for determining the sample size...

3. If the population size is around 500 (+/-100), approx 50% should be sampled.
4. If the population size is around 1500, approx 20% should be sampled.
5. Beyond a certain point (N = 5000), the population size is almost irrelevant and a sample size of 400 may be adequate.

Probability / Non Probability Sampling

Probability / Non Probability Sampling

- Probability sampling - the likelihood of any member of the population being selected is known

- Nonprobability sampling - the likelihood of any member of the population being selected is unknown

PROBABILITY SAMPLING STRATEGIES

- Simple random sampling
  - Each member of the population has an **equal** and **independent** chance of being chosen
  - The sample should be representative of the population

SYSTEMATIC SAMPLING

1. Jane	18. Steve	35. Fred
2. Bill	19. Sam	36. Mike
3. Harriet	20. Marvin	37. Doug
4. Leni	21. Ed. T.	38. Ed M.
5. Micah	22. Jerry	39. Tom
6. Sara	23. Chitra	40. Mike G.
7. Terri	24. Clenna	41. Nathan
8. Joan	25. Misty	42. Peggy
9. Jim	26. Cindy	43. Heather
10. Terrill	27. Sy	44. Debbie
11. Susie	28. Phyllis	45. Cheryl
12. Nona	29. Jerry	46. Wes
13. Doug	30. Harry	47. Genna
14. John S.	31. Dana	48. Ellie
15. Bruce A.	32. Bruce M.	49. Alex
16. Larry	33. Daphne	50. John D.
17. Bob	34. Phil	

1. Divide the population by the size of the desired sample: e.g., 50/10 = 5
2. Select a starting point at random: e.g., 43 = Heather
3. Select every 5th name from the starting point

STRATIFIED SAMPLING

- The goal of sampling is to select a sample that is representative of the population

- But suppose—
  - That people in the population differ systematically along some characteristic?
  - And this characteristic relates to the factors being studied?
- Then stratified sampling is one solution

STRATIFIED SAMPLING

- **Stratified sampling** is commonly used probability method that is superior to random sampling because it reduces sampling error. A stratum is a subset of the population that share at least one common characteristic.

- Examples of stratums might be males and females, or managers and non-managers. The researcher first identifies the relevant stratums and their actual representation in the population. Ranto a sadom sampling is then used to select a *sufficient* number of subjects from each stratum. "*Sufficient*" refers mple size large enough for us to be reasonably confident that the stratum represents the population. Stratified sampling is often used when one or more of the stratums in the population have a low incidence relative to the other stratums.

- The characteristic(s) of interest are identified (e.g., gender)

- The individuals in the population are listed separately according to their classification (e.g., females and males)

- The proportional representation of each class is determined (e.g., 40% females & 60% males)

- A random sample is selected that reflects the proportions in the population, (e.g., 4 females & 6 males)

STRATIFICATION ON MORE THAN ONE FACTOR
Hotel Clients

Location	Customer			Total
	Single	Couple	Family	
Local	1,200 [120]	1,200 [120]	600 [60]	3,000 [300]
National	2,800 [280]	2,800 [280]	1,400 [140]	7,000 [700]
International	4,000 [400]	4,000 [400]	2,000 [200]	10,000 [1000]

NONPROBABILITY SAMPLING STRATEGIES

- Convenience sampling
  - Captive or easily sampled population (20 people in your workplace)
  - Not random
  - Weak representativeness

- Quota sampling
  - Participants with the characteristic of interest are non-randomly selected until a set quota is met
- If you have been approached by a Researcher in the street, chances are this is Quota sampling.

- E.g. they may be doing Research for a new coffee shop. You may employ two different Quota controls, Age and Gender. Suppose then you have three age categories 16-34, 35 – 49 and 50 – 65 and that nationally, 70% of all coffee shop customers are in the 16 – 34 age group, 20% are in the 35-49 age group and 10% are in the 50-65 age group. You may also know that 60% are Female and 40% are Male. These guidelines would provide the quotas of people you must survey in order to ensure that the correct proportion of people within each category is represented in the sample.

- It could lead to bias – as you are only interviewing people who frequent the main street at the time of the day you are doing your research.

Summary
Sampling – Potential Problems

- Inappropriate Sampling Method
  - Probability Sampling
  - Non-Probability Sampling
- Inappropriate Sample Size
  - Too Large
  - Too Small
- Inappropriate Research Instrument
  - Questionnaires Vs Personal Interviews
  - Quantitative Vs Qualitative data

Quiz – Random Samples

The plant manager wants to study the process's ability to fill all bottles with 16 fluid ounces of cola. Why should she ensure that her sample data are random?

A. A random sample will allow her to draw conclusions about the bottles in her sample.

B. A random sample will allow her to draw conclusions about the entire population of filled bottles.

C. A random sample will not be representative of the entire population.

D. A random sample will be biased toward bottles filled with the target volume of fluid.

Quiz - Sampling

What should the automotive plant manager do to ensure that the sample of employees is representative of the population?

A. Ensure that assemblies don't exceed time limits.

B. Collect a sample of the first 10 available employees who assemble carburetors.

C. Sample employees who assemble carburetors the fastest.

D. Collect a random sample of employees from the population of employees who assemble carburetors.

End of Lecture