



OKAN ÜNİVERSİTESİ
İSTANBUL

BBA182 Applied Statistics

Week 1 (2) Introduction to Statistics

DR SUSANNE HANSEN SARAL

EMAIL: SUSANNE.SARAL@OKAN.EDU.TR

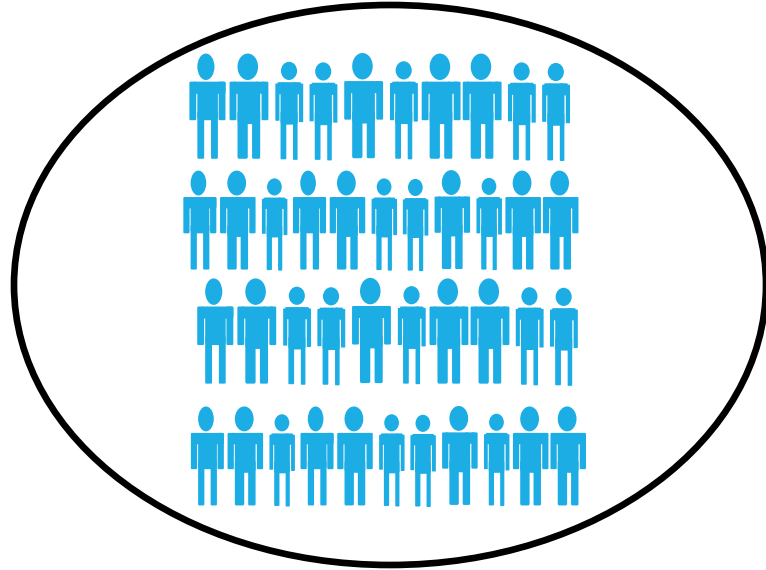
[HTTPS://PIAZZA.COM/CLASS/IXRJ5MMOX1U2T8?CID=4#](https://PIAZZA.COM/CLASS/IXRJ5MMOX1U2T8?CID=4#)

WWW.KHANACADEMY.ORG

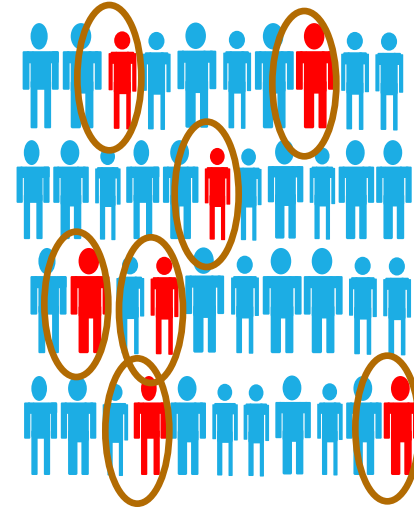


Population vs. Sample

Population



Sample





Statistical key definitions

POPULATION

A **population** is the collection of all items of interest under investigation. **N** represents the population size

Populations are usually very large, therefore it is impossible to investigate entire populations. It would be too

- Time consuming
- Costly



OKAN ÜNİVERSİTESİ
İSTANBUL

Statistical key definitions

SAMPLE

A **sample** is an observed subset of the population

- **n** represents the sample size



Statistical key definitions

PARAMETER VS. STATISTICS

A **parameter** is a specific characteristic of a population (mean, median, range, etc.)

Example: The **mean** (average) age of all students at OKAN

A **statistic** is a specific characteristic of a sample (sample mean, sample median, sample range, etc.)

Example: The **mean** (average) age of a sample of 500 students at OKAN



Why do we collect samples instead of investigating the entire population?

- Populations usually are infinite and their parameters are **rarely known.**
- The only way we can find the estimated value of a population parameter is by collecting a sample from the population of interest.



OKAN ÜNİVERSİTESİ
İSTANBUL

Why do we collect samples instead of investigating the entire population?

Populations are usually infinite. Therefore impossible to investigate the entire population

Less time consuming to investigate a **subset (sample)** of the population than investigating the entire population. Timely delivery of the results.

Less costly to administer, because workload is reduced

It is possible to obtain **statistical valid and reliable** results based on samples.



Randomness (Turkish: Rasgelelik)

Our final objective in statistics is to make **valid and reliable** statements about the population based on sample data. (inferential statistics)

Therefore we need a sample that represents **the entire** population

One important principle that we must follow in the sample selection process is **randomness**.



OKAN ÜNİVERSİTESİ
İSTANBUL

Main sampling techniques

Simple random sampling

Systematic sampling

Both techniques respect randomness and therefore provide **reliable** and **valid** data for statistical analysis



Random Sampling

Simple random sampling is a procedure in which:

- Each member/item in the population is chosen strictly by chance
- Each member/item in the population has an equal chance to be chosen
- Each member/item has to be independent from each other
- Every possible sample of n objects is equally likely to be chosen

The resulting sample is called a **random sample**.



Sampling error

In statistics we make decision about a population based on sample data, because the population parameter is unknown. Ex. Elections

Statisticians know that the sample statistic is rarely identical to the population parameter, but the two values are close.

The difference between the sample statistic and the population parameter is called **sampling error**.



OKAN ÜNİVERSİTESİ
İSTANBUL

Inferential statistics

Drawing conclusion about a population
based a sample information.



OKAN ÜNİVERSİTESİ
İSTANBUL

Inferential statistics

To draw conclusions about the population based on a sample we need to collect **data**.



What is data?

Data = information

Data can be **numbers**: Size of a hotel bill, number of hotel guests, number of nights stayed in a Hilton hotel, size of a swimming-pool, etc.

Data can be **categories**: Gender, Nationalities, marital status, tourist attractions, codes, university major, etc.



Data and context

Data are useless without a **context**.

When we deal with data we need to be able to answer at least the two following first questions in order to make sense of the data:

- 1) **Who?**
- 2) **What?**
- 2) When?
- 3) Where?
- 4) How?



Data and context

Data values are useless without their **context**

Consider the following:

Amazon.com may collect the following data:

10675489	Ohio	10.99	Chris G.
Samuel P.	10783489	Katherine H.	Canada
16.99	Monique D.	11.99	15783947
15.98	Massachusetts	12837593	Illinois

What information can we get out of this?



Data and context

We need to put the data into context in order to get information out of it

Purchase order #	Name	Price	Ship to State
10675489	Samuel P.	10.99	Ohio
10783489	Monique D.	16.99	Canada
15783947	Katherine H.	11.99	Massachusetts
12837593	Chris G.	15.98	Illinois



What is statistics?

It is a basic study of transforming data into information :

- how to collect it
- how to organize it
- how to summarize it, and finally
- to analyze and interpret it



Where does data come from?

- Market research
- Survey (online questionnaires, paper questionnaires, etc.)
- Interviews
- Research experiments (medicine, psychology, economics)
- Databases of companies, banks, insurance companies
- Internet
- other sources



Descriptive Statistics

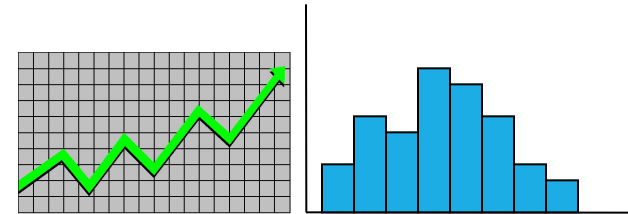
Collect data



- e.g., Survey, interview

Present data

- e.g., Tables and graphs



Summarize data

- e.g., Sample mean =

$$\frac{\sum X_i}{n}$$



OKAN ÜNİVERSİTESİ
İSTANBUL

Create your account in Khan Academy

Go to www.khanacademy.org create an account with your email address or your Facebook account (if you have one).

Add me (Susanne Hansen Saral) as a coach:

Follow the instructions from the hand-out



OKAN ÜNİVERSİTESİ
İSTANBUL

PIAZZA.COM

Piazza.com – class platform for:

Posting class lectures, course syllabus, class announcement, youtube videos, etc.