Math 243

Ch. 1: Data Collection

Statistics is the science of collecting, organizing, summarizing, and analyzing information to draw conclusions or answer questions. In addition, statistics is about providing a measure of confidence in any conclusions.

ank	Player	Country	Matches	Break points won
1	V.Azarenka	BLR	7	40
2	S.Williams	USA	7	33
3	S.Errani	ITA	6	30
3	M.Sharapova	RUS	6	30
5	R.Vinci	ITA	5	25
6	A.Ivanovic	SRB	5	24
6	M.Bartoli	FRA	5	24
8	A.Kerber	GER	4	23
9	A.Tatishvili	GEO	4	22
10	O.Puchkova	RUS	3	20
10	S.Stosur	AUS	5	20
12	A.Radwanska	POL	4	19
12	T.Pironkova	BUL	4	19
14	V.Lepchenko	USA	3	15
15	O.Govortsova	BLR	3	14
15	D.Cibulkova	SVK	3	14
15	S.Stephens	USA	3	14
15	K.Bertens	NED	2	14
15	A.Hlavackova	CZE	4	14
15	N.Petrova	RUS	4	14

The "Players"

The entire group of individuals to be studied is called the **population**.

A **sample** is a subset of the population that is being studied.

An **individual** is a person or object that is a member of the population being studied



Descriptive Statistics

Descriptive statistics consist of organizing and summarizing data. Descriptive statistics describe data through numerical summaries, tables, and graphs.

Example:

Descriptive statistics about a college involve the average math test score for incoming students. It says nothing about why the data is so or what trends we can see and follow.

A **statistic** is a numerical summary based on a sample.

Example :

You've performed a survey to 40 respondents about their favorite car color. And now you have a spreadsheet with the results.

However, this spreadsheet is not very informative and you want to summarize the data with some graphs and charts that can allow you to come up with some simple conclusions (e.g. 24% of people said that white is their favorite color).

Inferential statistics

Inferential Statistics uses methods that take results from a sample, extends them to the population, and measures the reliability of the result. Population



Topulation

Parameter vs. Statistic

A parameter is a numerical summary of a population.

- Suppose the percentage of all students on your campus who have a job is 84.9%. This value represents a **parameter** because it is a numerical summary of a population.
- Suppose a sample of 250 students is obtained, and from this sample we find that 86.3% have a job. This value represents a **statistic** because it is a numerical summary based on a sample.

Types of Data

Distinguish between Qualitative and Quantitative Variables: <u>Country</u> would be qualitative while <u>matches</u> is quantitative

WOMEN'S SINGLES BREAK POINTS WON LEADERS							
Rank	Player	Country	Matches	Break points won			
1	V.Azarenka	BLR	7	40			
2	S.Williams	USA	7	33			
3	S.Errani	ITA	6	30			
3	M.Sharapova	RUS	6	30			
5	R.Vinci	ITA	5	25			
6	A.Ivanovic	SRB	5	24			
6	M.Bartoli	FRA	5	24			
8	A.Kerber	GER	4	23			
9	A.Tatishvili	GEO	4	22			
10	O.Puchkova	RUS	3	20			
10	S.Stosur	AUS	5	20			
12	A.Radwanska	POL	4	19			
12	T.Pironkova	BUL	4	19			
14	V.Lepchenko	USA	3	15			
15	O.Govortsova	BLR	3	14			
15	D.Cibulkova	SVK	3	14			
15	S.Stephens	USA	3	14			
15	K.Bertens	NED	2	14			
15	A.Hlavackova	CZE	4	14			
15	N.Petrova	RUS	4	14			

Variable

Variables are the characteristics of the individuals within the population. It is what we are collecting Data about.

Examples: Break points won is a variable Matches played is a variable Nationality of the player is a variable

Types of Variables

- Qualitative or Categorical variables allow for classification of individuals based on some attribute or characteristic.
- **Quantitative variables** provide numerical measures of individuals. Arithmetic operations such as addition and subtraction can be performed on the values of the quantitative variable and provide meaningful results.

Example:

Researcher Elisabeth Kvaavik and others studied factors that affect the eating habits of adults in their mid-thirties. (Source: Kvaavik E, et. al. Psychological explanatorys of eating habits among adults in their mid-30's (2005) International Journal of Behavioral Nutrition and Physical Activity (2)9.) Classify each of the following variables considered in the study as qualitative or quantitative.

- a. Nationality
- b. Number of children
- c. Household income in the previous year
- d. Level of education
- e. Daily intake of whole grains (measured in grams per day)

Types of Quantitative Variables

- A discrete variable is a quantitative variable that either has a finite number of possible values or a countable number of possible values. The term "countable" means the values result from counting such as 0, 1, 2, 3, and so on.
- A **continuous variable** is a quantitative variable that has an infinite number of possible values it can take on and can be measured to any desired level of accuracy.





Example

Researcher Elisabeth Kvaavik and others studied factors that affect the eating habits of adults in their mid-thirties. (Source: Kvaavik E, et. al. Psychological explanatorys of eating habits among adults in their mid-30's (2005) International Journal of Behavioral Nutrition and Physical Activity (2)9.) Classify each of the following quantitative variables considered in the study as discrete or continuous.

- a. Number of children
- b. Household income in the previous year
- c. Daily intake of whole grains (measured in grams per day)