MATH 3339
Statistics for the Sciences

Wendy Wang
wwang60@central.uh.edu

Lecture 1

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Outline

1. Types of Data
2. Types of Variables
3. Sets
4. Venn Diagrams

Background

Chapter 1: Overview and Basic Concepts

What is Statistics?

- Statistics is used to make intelligent decisions in a world full of uncertainty. "A knowledge of statistics provides the necessary tool to differentiate between sound statistical conclusions and questionable conclusions." (*Business Statistics Communicating with Numbers*, Jaggia and Kelly, 2013, pg 4)
- Statistics is the science of collecting, organizing, and interpreting numerical facts which we call *data*.

A Data Set: Course Grades From Previous Session

variables

https://www.math.uh.edu/~cathy/Math3339/data/grades.txt

observation

Student	Score	Grade	Tests	Quiz	HW	Session
1	100.707	A	99.233	87.308	101.270	Sp16
2	81.310	B	75	98.231	64.444	Sp16
3	8.194	F	14.667	12.769	3.175	Sp16
4	90.449	A	91.533	77.231	82.222	Sp16
5	68.461	D	65.783	81.769	68.571	Sp16
6	103.955	A	103.32	97.923	101.905	Sp16
7	92.889	A	95.6	85.923	75.556	Sp16
8	84.805	B	83.2	79.385	75.238	Sp16
9	91.640	A	89.967	91.231	85.079	Sp16
10	22.316	F	17.433	40.615	44.444	Sp16
11	98.363	A	94.167	99.231	101.587	Sp16
12	49.250	F	43.917	73.077	78.095	Sp16
13	16.967	F	15.5	20.077	29.841	Sp16
14	50.747	F	45.533	67.385	57.460	Sp16
15	43.184	F	72.983	47.462	38.413	Sp16
16	100.845	A	98.667	96.231	100.317	Sp16
17	84.195	B	77.5	87.154	95.556	Sp16
18	84.400	B	78.733	78.615	82.540	Sp16
19	67.170	D	74.3	68.538	72.063	Fal15
20	87.413	B	92	82.077	77.778	Fal15
21	67.899	D	71.8	71.077	84.127	Fal15
22	74.676	C	70.083	83.308	73.016	Fal15
23	40.054	F	44.133	21.308	33.333	Fal15
24	101.014	A	101.08	98.923	95.873	Fal15
25	11.972	F	17.1	10.385	3.810	Fal15
26	79.831	B	86.233	71.923	46.667	Fal15
27	83.301	B	94.6	69.692	60.317	Fal15
28	72.299	C	64.967	67.615	99.394	Sum16
29	83.821	B	77.2	80.923	83.030	Sum16
30	90.703	A	83.617	87.923	80.000	Sum16

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Types of data

- **Population Data** is everything or everyone we want information about. It is a set of data that consists of all possible values pertaining to a certain set of observations or an investigation.

- **Sample Data** is a subset of the population that we have information from. It is just a small section of the population taken for the purpose of investigation.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Examples of Types of Data

Identify the population and the sample for each of the following:

- University of Houston is interested in how many students buy used books as opposed to new ones. They randomly choose 100 students at the student center to interview
  - Population - all UH students.
  - Sample - the 100 students interviewed.
- An elementary school is creating a new lunch menu. They send questionnaires to students with last names that begin with the letters M through R.
  - Population - all students in this elementary school
  - Sample - M - R.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Two Types of Variables

Go back to the example of **grades**. We have several variables, score, grade, tests, quiz, hw, & session.

- The variables **grade, & session** are *categorical variables*. **Categorical variables** place a case into one of several groups or categories.
- The variables **scores, tests, quiz & hw** are **quantitative variables**. **Quantitative Variables** take numerical values for which arithmetic operations such as adding and averaging make sense.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Two Types of Quantitative Variables

Quantitative variables can be classified as either **discrete** or **continuous**.

- Discrete quantitative variables - a countable set of values.
- Continuous quantitative variables - data that can take on any values within some interval.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Examples of Variables

Classify the following variables as categorical or quantitative. If quantitative, state whether the variable is discrete or continuous.

- Political preference.

  Categorical

- Number of siblings.

  quantitative — discrete

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Examples of Variables Part 2

Classify the following variables as categorical or quantitative. If quantitative, state whether the variable is discrete or continuous.

- Blood type.

  categorical

- Height of men on a professional basketball team.

  Quantitative — continuous.

- Time it takes to be on hold when calling the IRS at tax time.

  Quantitative $(0, \infty)$

Winning the State Lottery

- Suppose a person won the Jackpot of the State Lottery five times in a row.
- What do you think would happen?
- Is it possible for a person to win the state lottery five consecutive times?

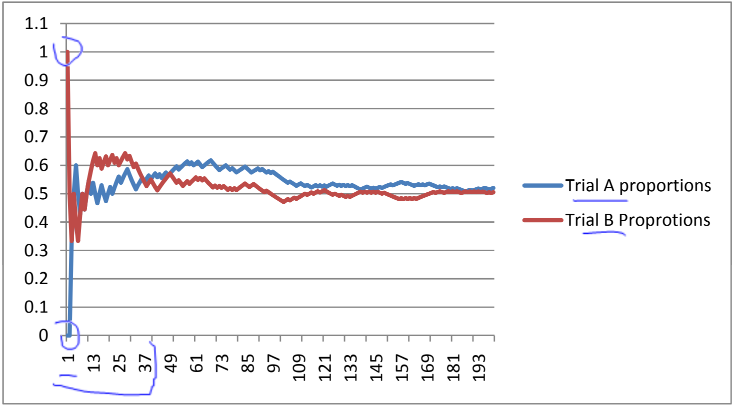
UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Randomness and Probability

- We call a phenomenon **random** if individual outcomes are uncertain.
- However, there is a regular distribution of outcomes in a large number of repetitions.
- Chance behaviors unpredictable in the short run but has a regular and predictable pattern in the long run.
- Long-run must be infinitely long to give them frequencies of enough time to even out.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Proportion of Heads in Long Run



UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Probability

- "A **probability** is a numerical value that measures the likelihood that an uncertain event occurs." *Business Statistics: Communicating with Numbers, Jaggia and Kelly, pg 96*
- The **probability** of any outcome of a random phenomenon is the proportion of times the outcome would occur in a very long series of repetitions.

Why Study Probability in Statistics?

- Statistical Idea: Rare event rule for inferential statistics
- If, under a given assumption, the probability of a particular observed event is extremely small, we conclude that the assumption is probably not correct.

< 0.0001

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Random Experiments

In order to study probability we need know how random experiments work. Examples of random experiments:

- Flipping a coin
- Asking who is the likely winner of a presidential election.
- Weighing Hershey chocolate bars.

A random experiment has the following two characteristics:

1. The experiment can be replicated an indefinite number of times under essentially the same experimental conditions.
2. There is a degree of uncertainty in the outcome of the experiment. The outcome may vary from replication to replications even though the experimental conditions are the same.

Sample Space

- A **set** is a collection of objects.
- The items that are in a set called **elements**.
- We typically denote a set by capital letters of the English alphabet. Usually, E_i
- Examples: $E_1 = \{knife, spoon, fork\}$, $E_2 = \{2, 4, 6, 8\}$.
- The set E_2 could also be written as
  $E_2 = \{x | x \text{ are even whole numbers between 0 and 10}\}$.
- The **sample space** of a random experiment is the set of all possible outcomes. Ω is used to denote sample space.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Notations of Sets

Notation	Description
$a \in A$	The object a is an element of the set A.
$A \subseteq B$	Set A is a subset of set B. That is every element in A is also in B.
$A \subset B$	Set A is a proper subset of set B. That is every element that is is in A is also in set B and there is at least one element in set B that is not in set A.
$A \cup B$	A set of all elements that are in "A **or** B." "A union B"
$A \cap B$	A set of all elements that are in "A **and** B." A & B
Ω	Called the **universal set**, all elements we are interested in.
$\sim A$ complement of A	The set of all elements that are in the universal set but are not in set A. $\bar{A}$ A^c A'
$\bigcup_i E_i$	$E_1 \cup E_2 \cup \ldots$, the union of multiple sets
$\bigcap_i E_i$	$E_1 \cap E_2 \cap \ldots$, the intersection of multiple sets

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Examples

The following are sets: $\Omega = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$,
$E_1 = \{1, 2, 3, 4, 5, 6, 9, 10\}$, $E_2 = \{3, 4, 7, 8\}$, and $E_3 = \{2, 3, 9, 10\}$

$$E_3 \subset E_1$$

$$E_1 \cap E_2 = \{3, 4\}$$

$$E_1 \cup E_2 = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$$
$$= \Omega$$

$$\sim E_3 = \overline{E_3} = E_3^c = \{1, 4, 5, 6, 7, 8\}$$

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Examples

The following are sets: $\Omega = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$,
$E_1 = \{1, 2, 3, 4, 5, 6, 9, 10\}$, $E_2 = \{3, 4, 7, 8\}$, and $E_3 = \{2, 3, 9, 10\}$

$$E_2 \sim E_1 = E_2 \cap \sim E_1 = E_2 \cap E_1^c$$

$$= \{3, 4, 7, 8\} \cap \{7, 8\}$$

$$= \{7, 8\}$$

$$E_1 \cap E_2 \cap E_3 = \{3\}$$

More Examples

The following are sets: $\Omega = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$, $E_1 = \{1, 2, 3, 4, 5, 6, 9, 10\}$, $E_2 = \{3, 4, 7, 8\}$, and $E_3 = \{2, 3, 9, 10\}$

1. What elements are in $\sim E_2 \cap \sim E_3$?

$\sim E_2 \cap \sim E_3 = \{1, 2, 5, 6, 9, 10\} \cap \{1, 4, 5, 6, 7, 8\}$

$= \{1, 5, 6\}$

2. What elements are in $E_2 \cap \sim E_1$?

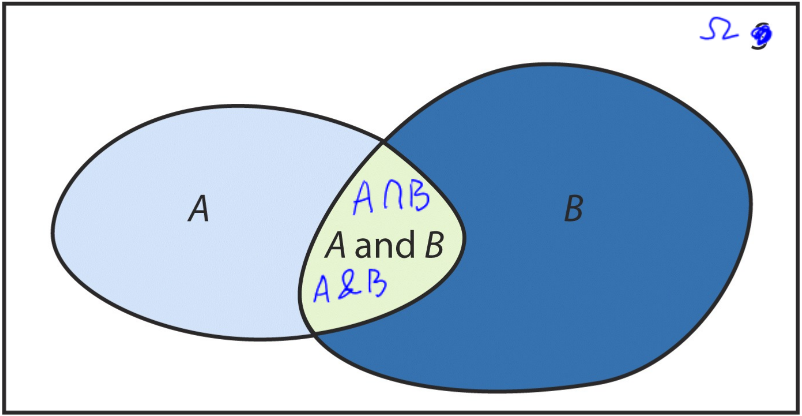
UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Definitions

- A **Venn diagram** is a very useful tool for showing the relationships between sets.
- Venn diagrams consist of a rectangle with one or more shapes (usually circles) inside the rectangle.
- The rectangle represents all of the elements that we are interested in for a given situation. This set is the universal set.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

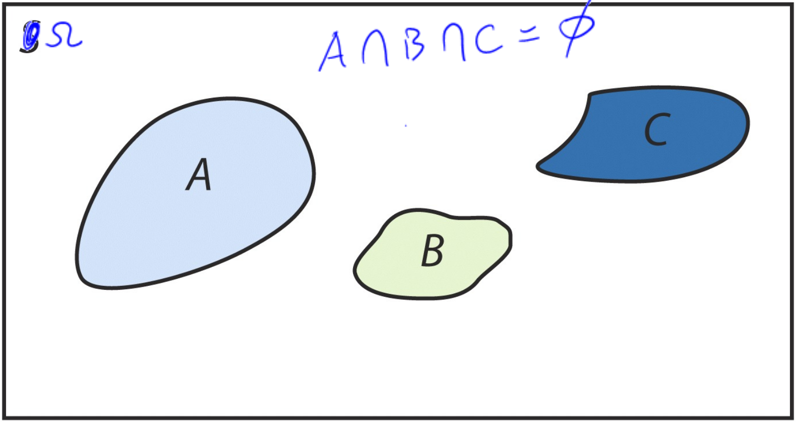
Graph of Venn Diagrams



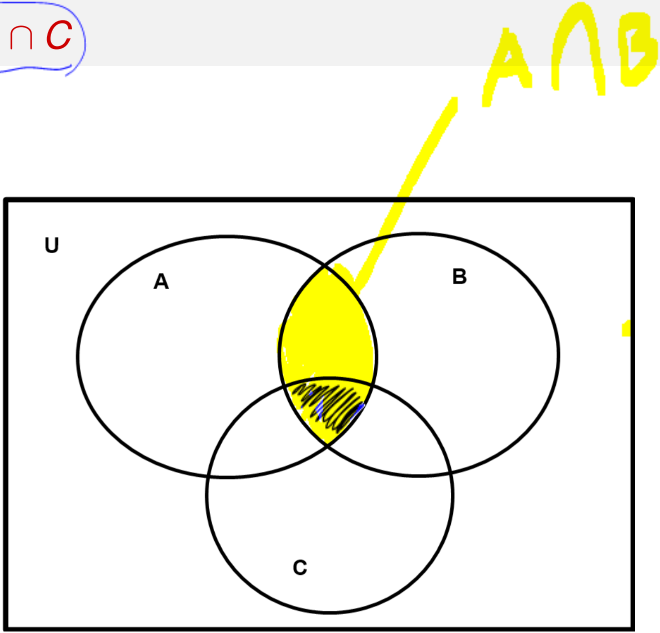
Graph of Disjoint Events

$A \& B = \phi$ $\quad$ $B \& C = \phi$

$A \cap B \cap C = \phi$



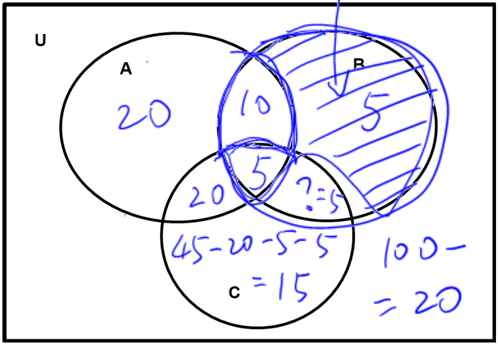
$A \cap B \cap C$



Soft Drink Preference

A group of 100 people are asked about their preference for soft drinks. The results are as follows: 55 like Coke, 25 like Diet Coke, 45 like Pepsi, 15 like Coke and Diet Coke, 5 like all 3 soft drinks, 25 like Coke and Pepsi, 5 only like Diet Coke (nothing else). Fill in the the Venn diagram with these numbers.

A: coke
B: Diet Coke
C: pepsi



UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS