

Definitions of Statistics, Probability, and Key Terms

Use the following information to answer the next five exercises. Studies are often done by pharmaceutical companies to determine the effectiveness of a treatment program. Suppose that a new AIDS antibody drug is currently under study. It is given to patients once the AIDS symptoms have revealed themselves. Of interest is the average (mean) length of time in months patients live once they start the treatment. Two researchers each follow a different set of 40 patients with AIDS from the start of treatment until their deaths. The following data (in months) are collected.

Researcher

A: 3; 4; 11; 15; 16; 17; 22; 44; 37; 16; 14; 24; 25; 15; 26; 27; 33; 29; 35; 44; 13; 21; 22; 10; 12; 8; 40; 32; 26; 27; 31; 34; 29; 17; 8; 24; 18; 47; 33; 34

Researcher

B: 3; 14; 11; 5; 16; 17; 28; 41; 31; 18; 14; 14; 26; 25; 21; 22; 31; 2; 35; 44; 23; 21; 21; 16; 12; 18; 41; 22; 16; 25; 33; 34; 29; 13; 18; 24; 23; 42; 33; 29

Determine what the key terms refer to in the example for Researcher A.

1. population
2. parameter
3. variable

Data, Sampling, and Variation in Data and Sampling

Use the following information to answer the next four exercises: A study was done to determine the age, number of times per week, and the duration (amount of time) of residents using a local park in San Antonio, Texas. The first house in the neighborhood around the park was selected randomly, and then the resident of every eighth house in the neighborhood around the park was interviewed.

The sampling method was

a. simple random; b. systematic; c. stratified; d. cluster

The colors of the houses around the park are what kind of data?

a. qualitative (categorical); b. quantitative discrete; c. quantitative continuous

The below table contains the total number of deaths worldwide as a result of earthquakes from 2000 to 2012.

Year	Total Number of Deaths
2000	231
2001	21,357
2002	11,685
2003	33,819
2004	228,802
2005	88,003
2006	6,605
2007	712
2008	88,011
2009	1,790
2010	320,120
2011	21,953

Year	Total Number of Deaths
2012	768
Total	823,856

Use this table to answer the following questions.

- What is the proportion of deaths between 2007 and 2012?
- What percent of deaths occurred before 2001?
- What is the percent of deaths that occurred in 2003 or after 2010?
- What is the fraction of deaths that happened before 2012?
- What kind of data is the number of deaths?
- Earthquakes are quantified according to the amount of energy they produce (examples are 2.1, 5.0, 6.7). What type of data is that?
- What contributed to the large number of deaths in 2010? In 2004? Explain.

For the following exercises, determine the type of sampling used (simple random, stratified, systematic, cluster, or convenience).

- A market researcher polls every tenth person who walks into a store.
- A computer generates 100 random numbers, and 100 people whose names correspond with the numbers on the list are chosen.

Use the following information to answer the exercises: Studies are often done by pharmaceutical companies to determine the effectiveness of a treatment program. Suppose that a new AIDS antibody drug is currently under study. It is given to patients once the AIDS symptoms have revealed themselves. Of interest is the average (mean) length of time in months patients live once starting the treatment. Two researchers each follow a different set of 40 AIDS patients from the start of treatment until their deaths. The following data (in months) are collected.

Researcher A: 3; 4; 11; 15; 16; 17; 22; 44; 37; 16; 14; 24; 25; 15; 26; 27; 33; 29; 35; 44; 13; 21; 22; 10; 12; 8; 40; 32; 26; 27; 31; 34; 29; 17; 8; 24; 18; 47; 33; 34

Researcher B: 3; 14; 11; 5; 16; 17; 28; 41; 31; 18; 14; 14; 26; 25; 21; 22; 31; 2; 35; 44; 23; 21; 21; 16; 12; 18; 41; 22; 16; 25; 33; 34; 29; 13; 18; 24; 23; 42; 33; 29

1. Determine what the key term data refers to in the above example for Researcher A.
2. Can you tell if one researcher is correct and the other one is incorrect? Why?
3. Suggest at least two methods the researchers might use to gather random data.

Suppose that the second researcher conducted his survey by choosing 40 patients he knew. What sampling method would that researcher have used? What concerns would you have about this data set, based upon the data collection method?

Use the following data to answer the exercises: Two researchers are gathering data on hours of video games played by school-aged children and young adults. They each randomly sample different groups of 150 students from the same school. They collect the following data.

Hours Played per Week	Frequency	Relative Frequency	Cumulative Relative Frequency
0–2	26	0.17	0.17
2–4	30	0.20	0.37
4–6	49	0.33	0.70
6–8	25	0.17	0.87
8–10	12	0.08	0.95
10–12	8	0.05	1

Table 1.29 Researcher A

Hours Played per Week	Frequency	Relative Frequency	Cumulative Relative Frequency
0–2	48	0.32	0.32
2–4	51	0.34	0.66
4–6	24	0.16	0.82
6–8	12	0.08	0.90
8–10	11	0.07	0.97
10–12	4	0.03	1

Table 1.30 Researcher B

1. Would the sample size be large enough if the population is the students in the school?
2. Researcher A concludes that most students play video games between four and six hours each week. Researcher B concludes that most students play video games between two and four hours each week. Who is correct?

As part of a way to reward students for participating in the survey, the researchers gave each student a gift card to a video game store. Would this affect the data if students knew about the award before the study?

Use the following data to answer the next five exercises: A pair of studies was performed to measure the effectiveness of a new software program designed to help stroke patients regain their problem-solving skills. Patients were asked to use the software program twice a day, once in the morning and once in the evening. The studies observed 200 stroke patients recovering over a period of several weeks. The first study collected the data in [Table 1.31](#). The second study collected the data in [Table 1.32](#).

Group	Showed improvement	No improvement	Deterioration
Used program	142	43	15

Group	Showed improvement	No improvement	Deterioration
Did not use program	72	110	18

Table 1.31

Group	Showed improvement	No improvement	Deterioration
Used program	105	74	19
Did not use program	89	99	12

Table 1.32

1. Given what you know, which study is correct?
2. Patients who used the software were also a part of an exercise program whereas patients who did not use the software were not. Does this change the validity of the conclusions from [Exercise 1.31](#)?
3. Is a sample of 500 volunteers a reliable measure for a population of 2,500?
4. Is a sample size of two representative of a population of five?

Frequency, Frequency Tables, and Levels of Measurement

What type of measure scale is being used? Nominal, ordinal, interval or ratio.

- a. High school soccer players classified by their athletic ability: Superior, Average, Above average
- b. Baking temperatures for various main dishes: 350, 400, 325, 250, 300
- c. The colors of crayons in a 24-crayon box
- d. Social security numbers
- e. Incomes measured in dollars
- f. A satisfaction survey of a social website by number: 1 = very satisfied, 2 = somewhat satisfied, 3 = not satisfied

- g. Political outlook: extreme left, left-of-center, right-of-center, extreme right
- h. Time of day on an analog watch
- i. The distance in miles to the closest grocery store
- j. The dates 1066, 1492, 1644, 1947, and 1944
- k. The heights of 21–65 year-old women
- l. Common letter grades: A, B, C, D, and F

Experimental Design and Ethics

Discuss potential violations of the rule requiring informed consent.

- a. Inmates in a correctional facility are offered good behavior credit in return for participation in a study.
- b. A research study is designed to investigate a new children's allergy medication.
- c. Participants in a study are told that the new medication being tested is highly promising, but they are not told that only a small portion of participants will receive the new medication. Others will receive placebo treatments and traditional treatments.