

Statistics

The study of data: how to collect, analyze, summarize, and present it.

Types of statistics:

1. Descriptive statistics
2. Inferential statistics

Descriptive statistics

- Descriptive statistics is mainly for understanding data, simply the analysis of data that helps describe, show or summarize data in a meaningful way.
- Descriptive statistics is a branch of statistics that deals with the collection of data to give a precise information about which data is obtained.

Types of Descriptive Statistics

1) Measures of Central Tendency – Mean, Median and Mode

Mean - The mean is also known as “M” and is the most common method for finding averages. You get the mean by adding all the response values together, dividing the sum by the number of responses, or “N.”

Mode - The mode is just the most frequent response value

Median - Finally, we have the median, defined as the value in the precise center of the dataset.

2) Measure of Variability (or dispersion)

- which shows how spread out the values are – Range, Variance and SD
- Dispersion refers to the spread of the values around the central tendency.

Range - Use range to determine how far apart the most extreme values are

SD - standard deviation (s) is your dataset's average amount of variability, showing you how far each score lies from the mean. Low standard deviation implies that most values

are close to the mean. High standard deviation suggests that the values are more broadly spread out. The larger your standard deviation, the greater your dataset's variable. Follow these six steps:

1. List the scores and their means.
2. Find the deviation by subtracting the mean from each score.
3. Square each deviation.
4. Total up all the squared deviations.
5. Divide the sum of the squared deviations by $N-1$.
6. Find the result's square root.

Raw Number/Data	Deviation from Mean	Deviation Squared
4	$4-7.3= -3.3$	10.89
6	$6-7.3= -1.3$	1.69
7	$7-7.3= -0.3$	0.09
8	$8-7.3= 0.7$	0.49
8	$8-7.3= 0.7$	0.49
9	$9-7.3=1.7$	2.89
10	$10-7.3= 2.7$	7.29
M=7.3	Sum = 0.9	Square sums= 23.83

When you divide the sum of the squared deviations by 6 ($N-1$): $23.83/6$, you get 3.971, and the square root of that result is 1.992. As a result, we now know that each score deviates from the mean by an average of 1.992 points.

Difference between Mean absolute deviation vs standard deviation

Variance - Variance reflects the dataset's degree spread. The greater the degree of data spread, the larger the variance relative to the mean.

Inferential statistics

- Inferential differs from descriptive in that inferential stats makes guesses about the population whereas descriptive describes the subset.
- Inferential statistics is where you use characteristics of a sample to infer something about a population.
- Inferential statistics can be defined as a field of statistics that uses analytical tools for drawing conclusions about a population by examining random samples. The goal of inferential statistics is to generalize about a population. In inferential statistics, a statistic is taken from the sample data (e.g., the sample mean) that used to make inferences about the population parameter (e.g., the population mean).

Types of inferential statistics

- 1) Hypothesis Testing
- 2) Regression Analysis

Hypothesis Testing:

Hypothesis testing is a type of inferential statistics that is used to test assumptions and draw conclusions about the population from the available sample data. It involves setting up a null hypothesis and an alternative hypothesis followed by conducting a statistical test of significance. A conclusion is drawn based on the value of the test statistic, the critical value, and the confidence intervals.

Regression Analysis:

- Regression analysis is used to quantify how one variable will change with respect to another variable.
- Checks the effect of a unit change of the independent variable in the dependent variable

Difference between Descriptive and inferential statistics:

- 1) Descriptive statistics doesn't infer any conclusions or predictions, which implies that inferential statistics do so.
- 2) Inferential statistics takes a random sample of data from a portion of the population and describes and makes inferences about the entire population. For instance, in asking 50 people if they liked the movie they had just seen, inferential statistics would build on that and assume that those results would hold for the rest of the moviegoing population in general.
- 3) Therefore, if you stood outside that movie theater and surveyed 50 people who had just seen Rocky 20: Enough Already! and 38 of them disliked it (about 76 percent), you could extrapolate that 76% of the rest of the movie-watching world will dislike it too, even though you haven't the means, time, and opportunity to ask all those people.
- 4) Simply put: Descriptive statistics give you a clear picture of what your current data shows. Inferential statistics makes projections based on that data.

	Descriptive Statistics	Inferential Statistics
1	It gives information about raw data which describes the data in some manner.	It makes inference about population using data drawn from the population.
2	It helps in organizing, analyzing and to present data in a meaningful manner.	It allows us to compare data, make hypothesis and predictions.
3	It is used to describe a situation.	It is used to explain the chance of occurrence of an event.
4	It explains already known data and limited to a sample or population having small size.	It attempts to reach the conclusion about the population.

Types of sampling:

- **Population:** The whole group we are interested in
- **Census:** A collection of data from the whole population
- **Sample:** A collection of data from part of the population

But how do we choose what members of the population to sample?

There are 4 main methods:

1) Random Sampling

- The best way is to choose randomly
- Imagine slips of paper each with a person's name, put all the slips into a barrel, mix them up, then dive your hand in and choose some slips of paper.
- But this means you need a full list of the population to choose from.

Example: You want to know the favorite colors for people at your school, but do not have the time to ask everyone.

- Somehow get a full list of students printed out, then place all pages on the ground, drop a pencil and note down the student's name.
- Repeat until you have 50 names. Now survey those 50. Your results will hopefully be nearly as good as if you had asked everyone.
- Random surveys are the best way to avoid bias.
- And your results are better when you ask more people.

Example: nationwide opinion polls survey around 2,000 people, and the results are nearly as good (within about 1%) as asking everyone.

2) Systematic Sampling

- This is where we follow some system of selection like "every 10th person"

Example: You want to know the favorite colors for people at your school, but don't have the time to ask everyone.

Solution: stand at the gate and choose "every 4th person to arrive"

3) Stratified Sampling

- This is where we divide the population into groups by some characteristic such as age or occupation or gender.
- Then make sure our survey includes people from each group in proportion to how many there are in the whole population.

Example: We want to survey 300 people in the USA

This is the population breakdown for the USA in 2010:

Age Range	Percent
0-4	6.50%
17-May	17.50%
18-23	9.90%
24-44	26.60%
45-64	26.40%
65+	13.00%
	100%

We want to survey 300 people, so we choose:

Age Range	Percent	People
0-4	6.50%	20
17-May	17.50%	52
18-23	9.90%	30
24-44	26.70%	80
45-64	26.40%	79
65+	13.00%	39
	100%	300

4) Cluster Sampling

- We break the population into many groups, then randomly choose whole groups.
- Example: we divide the town into many different zones, then randomly choose 5 zones and survey everyone in those zones.
- Cluster sampling works best when the clusters are similar in character to each other.

Example: if the town has rich and poor zones then try to create a new way of dividing the town into fairer regions.

Quiz

1. Mrs. Trahan samples her class by selecting 5 girls and 7 boys. This type of sampling is called.
2. Mrs. Trahan samples her class by selecting every third person on her class list. Which type of sampling method is this?
3. Mrs. Trahan samples her class by selecting all students sitting at group 1 and group 5 in her classroom. This sampling technique is called.
4. Mrs. Trahan samples her class by picking 10 numbers from her hat and each number is assigned to a student.
5. In order to use samples to estimate something from the population, the sample should be _____ the population.
6. If every individual in a population has the same chance of being included in a sample, the sample is a(n) _____ sample.
7. The school librarian wants to determine how many students use the library on a regular basis. What type of sampling method would she use if she chose to randomly select 20 students in the cafeteria during each of the three lunch periods on Monday.
8. A student wants to determine the favorite professional sport of students in her high school. Which of the following samples is most likely to give her a representative sample?

answer choices

A random Sample of the students in the art honor society

A random sample of students on the basketball team

A random sample of students on the official school enrollment roster

A random sample of students in the library during lunch

9. To determine the most popular type of cake amount all the students at your school, ask your entire math class.