# Tutorial 1 Introduction to Statistics

1. Indicate whether each of the following examples refers to a population or to a sample.
2. A group of 25 patients selected to test a new drug
3. Total items produced on a machine for each year from 2014 to 2016
4. Yearly expenditures on clothes for 50 persons
5. Number of houses sold by each of the 10 employees of a real estate agency during 2016
6. Salaries of CEOs of all companies in New York City
7. Five hundred houses selected from a city
8. Gross sales for 2016 of four fast-food chains
9. Annual incomes of all 33 employees of a restaurant
10. Classify the following data as discrete data or continuous data.
11. Time taken to complete a 100 meters run.
12. The number of cars parked in this parking lot.
13. Diameter of the basketballs manufactured by Company A.
14. The width of the rooms in the campus.
15. The size of the shoes sold in a day.
16. The following table gives the names, gender and salary for the past week for five workers.

|  |  |  |
| --- | --- | --- |
| Name | Gender | Salary |
| Ali | Male | 420 |
| David | Male | 330 |
| John | Male | 280 |
| Kathy | Female | 470 |
| Steven | Male | 400 |

1. What are the variables in this data set? Indicate whether they are quantitative or qualitative.
2. How many elements does this data set contain? Give 2 of the elements.
3. The number of restaurants in each of five small towns is 4, 12, 8, 10 and 5, respectively. Let *x* denote the number of restaurants in a small town. Find:
4.  **b.**  **c.** 
5. The following table lists five pairs of *m* and *f* values.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *m* | 3 | 16 | 11 | 9 | 20 |
| *f* | 7 | 32 | 17 | 12 | 34 |

Compute the value of each of the following:

**a.**  **b.**  **c.**  **d.**  **e.** 

1. The number of passengers in 20 automobiles taxi on a highway is

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 2 | 2 | 4 | 3 | 1 | 1 | 5 | 0 | 1 | 3 |
| 3 | 1 | 4 | 0 | 1 | 1 | 3 | 0 | 1 | 2 |

1. Construct a frequency table.
2. Find the mean, mode and median for the data.
3. The following table gives the scores of fifty students on a statistics test.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 73 | 65 | 82 | 70 | 45 | 50 | 70 | 54 | 32 | 75 |
| 75 | 67 | 65 | 60 | 75 | 87 | 83 | 40 | 72 | 64 |
| 58 | 75 | 89 | 70 | 73 | 55 | 61 | 78 | 89 | 93 |
| 43 | 51 | 59 | 38 | 65 | 71 | 75 | 85 | 65 | 85 |
| 49 | 97 | 55 | 60 | 76 | 75 | 69 | 35 | 45 | 63 |

* 1. Prepare a stem-and-leaf display of these values.
  2. Construct a frequency table.
  3. Using the frequency table from part (b), compute the mean, variance and standard deviation for the data.

1. A network provider investigates the load of its network. The number of concurrent users is recorded at fifty locations (thousands of people),

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 17.2 | 22.1 | 18.5 | 17.2 | 18.6 | 14.8 | 21.7 | 15.8 | 16.3 | 22.8 |
| 24.1 | 13.3 | 16.2 | 17.5 | 19 | 23.9 | 14.8 | 22.2 | 21.7 | 20.7 |
| 13.5 | 15.8 | 13.1 | 16.1 | 21.9 | 23.9 | 19.3 | 12 | 19.9 | 19.4 |
| 15.4 | 16.7 | 19.5 | 16.2 | 16.9 | 17.1 | 20.2 | 13.4 | 19.8 | 17.7 |
| 19.7 | 18.7 | 17.6 | 15.9 | 15.2 | 17.1 | 15 | 18.8 | 21.6 | 11.9 |

1. Compute the sample mean, variance, and standard deviation of the number of concurrent users.
2. Compute the interquartile range. Are there any outliers?
3. The mean time taken to learn the basics of a word processor by all students is 200 minutes with a standard deviation of 20 minutes.
4. Using Chebyshev’s theorem, find at least what percentage of students will learn the basics of this word processor in 160 to 240 minutes.
5. Using Chebyshev’s theorem, find the interval that contains the time taken by at least 84% of all students to learn this word processor.

Suppose the times taken to learn the basics of this word processor by all students now have a bell-shaped distribution with a mean of 200 minutes and a standard deviation of 20 minutes.

1. Using the empirical rule, find the percentage of students who will learn the basics of this word processor in180 to 220 minutes.
2. Using the empirical rule, find the interval that contains the time taken by 99.7% of all students to learn this word processor.
3. Consider three data sets.

(1)

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 19 | 24 | 12 | 19 | 18 | 24 | 8 | 5 | 9 | 20 |
| 13 | 21 | 21 | 20 | 19 | 15 | 15 | 26 | 16 | 1 |
| 13 | 11 | 1 | 12 | 11 | 10 | 22 | 21 | 7 | 16 |

(2)

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 17 | 24 | 21 | 22 | 26 | 22 | 19 | 21 | 23 | 11 |
| 18 | 16 | 20 | 21 | 20 | 23 | 33 | 19 | 21 | 24 |
| 19 | 14 | 23 | 25 | 26 | 15 | 17 | 26 | 21 | 18 |

(3)

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 56 | 52 | 13 | 34 | 33 | 18 | 44 | 41 | 48 | 75 |
| 39 | 53 | 23 | 41 | 78 | 15 | 35 | 40 | 18 | 15 |
| 24 | 19 | 35 | 27 | 46 | 62 | 71 | 24 | 66 | 94 |

1. For each data set, draw a histogram and determine whether the distribution is right skewed, left-skewed, or symmetric.
2. Compute sample means and sample medians. Do they support your findings about skewness and symmetry? How?
3. Given a person jogged for a total of 46, 50, 52 and 60 minutes on four Saturdays.
4. Find the mean, range and variance of these four values.
5. Subtract 50 minutes from each of the times, recalculate the mean, range and variance, and compare the results with those obtained in part (a).
6. Divide each of the original values by 2, recalculate the mean, range and variance, and compare the results with those obtained in part (a).
7. The following data set represents the number of all new computer accounts registered during ten consecutive days.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 43 | 37 | 50 | 51 | 58 | 105 | 52 | 45 | 45 | 10 |

1. Compute the mean, median, quartiles, and standard deviation.
2. Check for outliers using the 1.5(IQR) rule.
3. Delete the detected outliers and compute the mean, median, quartiles, and standard deviation again.
4. Make a conclusion about the effect of outliers on basic descriptive statistics.

Answers to selected exercises.

4. (a) 39 (b) 1521 (c) 349

5. (a) 56 (b) 102 (c) 1508 (d) 867 (e) 24884

6. (b) 1.9, 1, 1.5

8. (a) 17.95, 9.97, 3.16 (b) 4.1, no outlier

9. (a) 75% (b) (150, 250) (c) 68% (d) (140, 260)

10. (b) (1) 14.97, 15.5 (2) 20.83, 21 (3) 41.3, 39.5

11. (a) 52, 14, 5.89

12. (a) 49.6, 47.5, 43, 52, 22.27 (b) 13.5 (c) 47.65, 47.5, 44, 51.5, 6.04