STATISTICS FOR GEOLOGY COURSE

G230

By Alaa M. Atiaa

Ph. D.

LESSON TWO

Frequency Distributions and Graphs

Outline

2-1   Introduction

2-2   Organizing Data

2-3   Histogram, Frequency Polygons, and Ogives

2-4   Other Types of Graphs

2-1   Introduction

When conducting a statistical study, the researcher must gather data for the variable under study. For example, if a researcher wishes to study the number of earthquake in a specific geographic area over the past several years. He or she has to gather data from various seismic stations within the area.

To describe situations, draw conclusions, or make inference, the research must organize the data in some meaningful way. The most convenient method of organizing data is to conduct a frequency distribution.

A **frequency distribution** is the organization of row data in table form, using classes and frequencies

Categorical Frequency Distribution

The categorical frequency distribution is used for data that can be placed in specific categories, such as nominal or ordinal level data. For example, data such as political affiliation, religious affiliation, or major field of study would use categorical frequency distributions.

Example 2-2

25 army inductees were given a blood test to determine their blood type. The data set is

A	B	B	AB	O
O	O	B	AB	B
B	B	O	A	O
A	O	O	O	AB
AB	A	O	B	A

Construct a frequency distribution for the data.

Solution

Since the data are categorical, discrete classes can be used. There are four blood types: A, B, O, and AB.

These types will be used as the classes for the distribution.

The frequency table for this example is shown below

Class limits	Frequency	Percent
A	5	20
B	7	28
O	9	36
AB	4	16
	Total = 25	100

General outlines to construct frequency table

STEP 1   Determine the classes

Find the highest and lowest value

Find the range (Range = maximum value – minimum value)

Select the number of classes desired (use the following equation)

$$No. of\ classes = 1 + (3.3 \times \log N)$$

Find the width by dividing the range by the number of classes and rounding up.

$$class\ width = \frac{Range}{No.\,of\,classes}$$

Select a starting point (usually the lowest value or any convenient number less than the lowest value); add the width to get lower limits.

STEP 2   **Numerate the data**

STEP 3   **Find the numerical frequencies**

STEP 4   **Find the cumulative frequencies**

Example 2-2

These data represent the record high temperature for each of the 50 states. Construct a frequency table for the data.

112	100	127	120	134	118	105	110	109	112
110	118	117	116	118	122	114	114	105	109
107	112	114	115	118	117	118	122	102	110
116	108	110	121	113	120	119	111	104	111
120	113	120	117	105	110	118	112	114	114

Solution

---

1. Find the lowest (L) and highest (H) values

L = 100, H = 134

2. Find Range (R = 134 – 100 = 34)

3. Find Number of classes ($No.\,of\ classes = 1 + (3.3 \times \log N)) = 1 + (3.3 * \log 50) = 6.6$

   Rounded up 6.6 = **7**

4. Find width of class

$$class\ width = \frac{Range}{No.\,of\,classes} = \frac{34}{7} = 4.9$$

   Rounded up 4.9 = 5

5. Select a starting point for the lowest class limit. This can be the smallest data value or any convenient number less than the smallest data value. In this case, 100 is used. Add the width to the lowest lower score taken as the starting point to get the lower limit of the next classes. Keep adding until there are 7 classes as shown 100, 105, 110, etc.

6. Find the numerical frequencies for each class.
7. Find the cumulative frequencies.

The completed frequency distribution is:

Class limits	Frequency	Cumulative frequency
100 – 105	6	6
105 – 110	9	15
110 – 115	14	29
115 – 120	16	45
120 – 125	3	48
125 – 130	1	49
130 – 135	1	50

Geological Example

Example 2-3

The following data for the pay-zone thickness (in feet) are collected from all available wells in a reservoir:

6	32	23
10	27	36
20	10	19
12	18	13
20	29	33
10	8	10
15	17	26

Analyze this data values using a frequency-distribution analysis.

Solution

1. Find the lowest (L) and highest (H) values

L = 6, H = 36

2. Find Range (R = 36 − 6 = **30**)

3. Find Number of classes ($No. of\ classes = 1 + (3.3 \times \log N)) = 1 + (3.3 * \log(\mathbf{21})) = 5.36$

   Rounded up 5.36 = **6**

4. Find width of class

$$class\ width = \frac{Range}{No. of classes} = \frac{30}{6} = 5$$

   It does not need rounded up.

5. Select a starting point for the lowest class limit. This can be the smallest data value or any convenient number less than the smallest data value. In this case, 6 is used. Add the width to the lowest lower score taken as the starting point to get the lower limit of the next classes. Keep adding until there are 6 classes are created.

6. Find the numerical frequencies for each class.

7. Find the cumulative frequencies.

The completed frequency distribution is:

Class limits	Frequency	Cumulative frequency
6 – 11	6	6
11 – 16	3	9
16 – 21	5	14
21 - 26	2	16
26 – 31	2	18
31 – 36	3	21

Exercise 2.1

The following data represent the concentration of Magnesium in ppm in soil samples collected in vineyards associated terraces on the Istrian peninsula of Croatia, analysis them using frequency distribution technique.

6936	7450	2000	3100
5368	4400	1000	7100
5006	5000	3100	6400
3600	8600	5000	7900
3220	4000	9000	2300

Histogram, Frequency polygons, and Ogives

After the data have been organized into a frequency distribution, they can be presented in graphical form. The purpose of graphs in statistics is to convey the data to the viewers in pictorial form. It is easier for most people to comprehend the meaning of data presented graphically than data presented numerically in tables or frequency distributions.

The three most commonly used graphs in research are

1. The histogram
2. The frequency polygon
3. The cumulative frequency graph, or ogive

1. The histogram

The following steps explain how to draw a histogram

1. From the frequency table using the class limits as the x-axis and the frequencies on the y-axis.
2. Using the frequencies as the heights and draw vertical bars for each class.

Example 2.4

For the following row data, make a frequency table and draw histogram.

| 6	32	23
10	27	36
20	10	19
12	18	13
20	29	33
10	8	10
15	17	26

Solution

The frequency table for these data is already prepared in the example 2.3. We added the calculations of midpoint values to draw the other graphs as well.

This is the frequency table of the above data

| Class limits	Class midpoint	Frequency	Cumulative frequency
6 – 11	8.5	6	6
11 – 16	13.5	3	9
16 – 21	18.5	5	14
21 - 26	23.5	2	16
26 – 31	28.5	2	18
31 – 36	33.5	3	21

The histogram of this table is as explained below
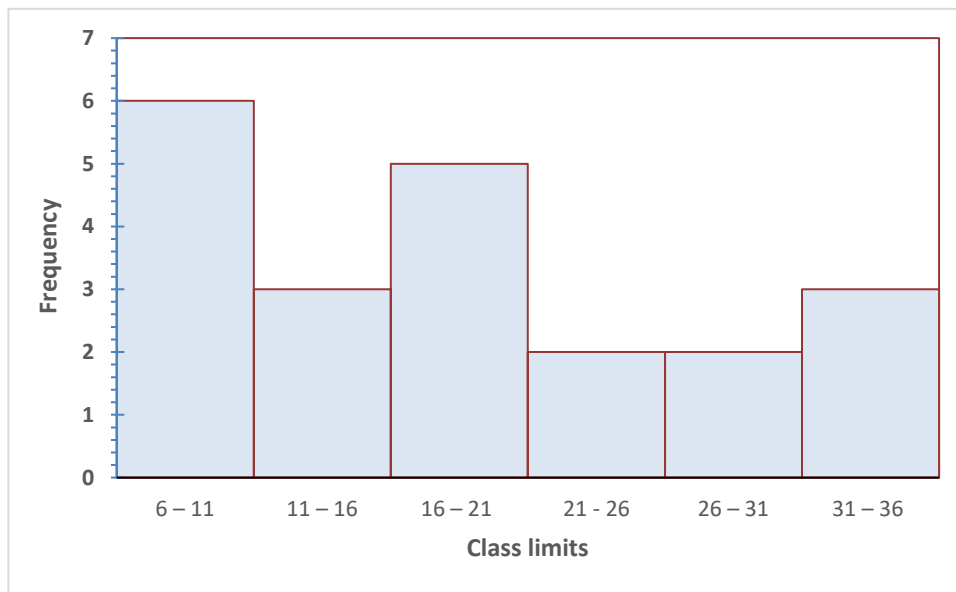
2. The frequency polygon

The frequency polygon is a graph that displays the data by using lines that connect points plotted for the frequencies at the midpoints of the classes. The frequencies are represented by the heights of the points. The midpoint of each class represents the x-axis while the y-axis represents the frequencies. The midpoints are found by adding the upper and lower classes limits and dividing by 2 (ie the average).

For the frequency table in the previous example, the frequency polygon is as follows:
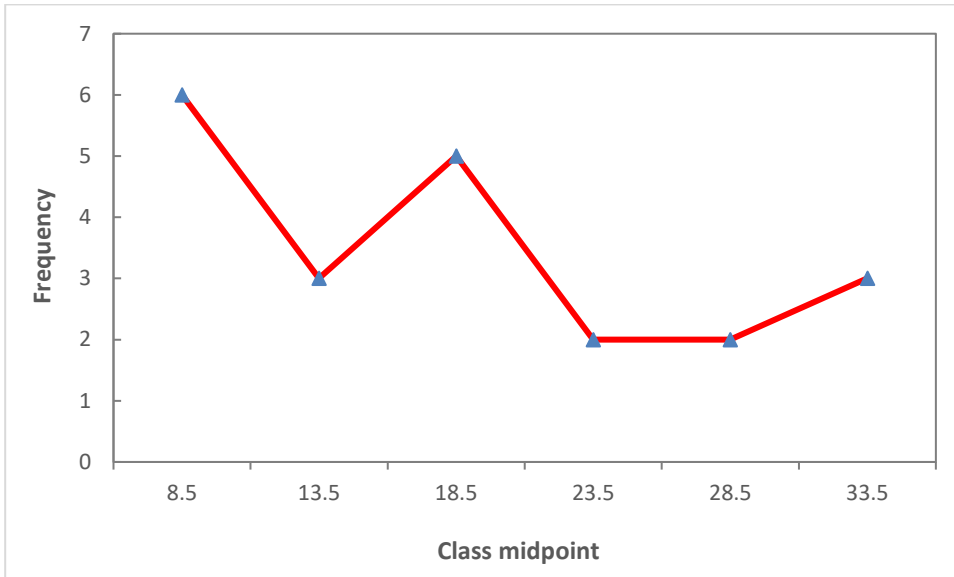
3. The ogive

To draw the ogive, the cumulative frequency is plotted on the y-axis and the class limits is on the x-axis as line type plot.

For the previous example, the following ogive plot is obtained.