Tables and Graphical presentations
Lecture 2

Lecture objectives

- Identify the two variable types
- Identify measurement level for each variable
- Organize data using a frequency distribution.

1 Variable

A variable is a characteristic or attribute that can assume different values.
If the values are determined by chance it called *random variable.*

Variables can be classified as:

1.1 Qualitative Variables

Variables which assume non-numerical values.

1.2 Quantitative Variables

Variables which assume numerical values. Quantitative variables can be further classified into two types:

1.2.1 Discrete Variables

Variables which assume a finite or countable number of possible values. Usually obtained by counting, for example number of children in a family or the grade of student.

1.2.2 Continuous Variables

Variables which assume an infinite number of possible values. Usually obtained by measurement. They often include fractions and decimals, for example temperature and weight.

2 Measurement Levels

In addition to being classified as qualitative or quantitative, variables can be classified by how they measured. There are four levels of measurement: Nominal, Ordinal, Interval, and Ratio.

2.1 The nominal level of measurement

classifies data into mutually exclusive (non overlapping) categories in which no order or ranking can be imposed on the data, for example gender

2.2 The ordinal level of measurement

classifies data into categories that can be ranked; however, precise differences between the ranks do not exist, for example student grade (A, B, C,D and F)

2.3 The interval level of measurement

ranks data, and precise differences between units of measure do exist; however, there is no meaningful zero, for example temperature and scale IQ

2.4 The ratio level of measurement

possesses all the characteristics of interval measurement, and there exists a true zero. In addition, true ratios exist when the same variable is measured on two different members of the population, for example height, weight.

3 Frequency Distribution

When conducting a statistical study, the researcher must gather data for the particular variable under study. For example, if a researcher wishes to study the number of people who were been into a car accident in Khartoum State in the past year, he or she has to gather the data from various doctors, hospitals, or health departments. To describe situations, draw conclusions, or make inferences about events, the researcher must organize the data in some meaningful way. The most convenient method of organizing data is to construct *frequency distribution*

3.1 Frequency Distribution definition

Frequency distribution is the organization of raw data in table form, using classes and frequencies.

3.2 Raw Data

Is the collected data in the original form.

There are types of frequency tables categorical frequency distribution, grouped frequency distribution and ungrouped frequency distribution.

3.2.1 Categorical Frequency Table

is used for data that can be placed in specific categories, such as nominal or ordinal level data. For example blood types, student grades. We can add Relative frequency. **Relative frequency** is the frequency divided by the total frequency. This gives the percent of values falling in that class.

Example 1 *twenty five patients were given a blood test to determine their blood type. The data set is*

Construct a frequency table for the data

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

Solution

First Make a table with the blood types in the first column
Second Tally the data and place the results in the tally column.
Third Count the tallies and place the results in the frequencies column.
Fourth Find the percentage of values in each class using

$$percent = \frac{freq}{total}$$

the resulting table will be like

class	tally	frequency	percent
A	$\not{\|\|\|\|}$	5	20%
B	$\not{\|\|\|\|}$ $\|\|$	7	28%
O	$\not{\|\|\|\|}$ $\|\|\|$	9	36%
AB	$\|\|\|$	4	16%
		Total=25	100%

As a result form the table more people have type O blood than any other type.

3.2.2 Grouped Frequency Distribution

When the range of the data is large, the data must be grouped into classes that are more than one unit in width, for example the number of hours spent online.

There are some concepts must be study before the example:

Class Limits Separate one class in a grouped frequency distribution from another. The limits could actually appear in the data and have gaps between the upper limit of one class and the lower limit of the next.
Class Boundaries Separate one class in a grouped frequency distribution from another. The boundaries have one more decimal place than the raw data and therefore do not appear in the data. There is no gap between the upper boundary of one class and the lower boundary of the next class. The lower class boundary is found by subtracting 0.5 units from the lower class limit and the upper class boundary is found by adding 0.5 units to the upper class limit.
Class Width The difference between the upper and lower boundaries of any class. The class width is also the difference between the lower limits of two consecutive classes or the upper limits of two consecutive classes. It is not the difference between the upper and lower limits of the same class.
Class Mark (Midpoint) The number in the middle of the class. It is found by adding the upper and lower limits and dividing by two. It can also be found by adding the upper and lower boundaries and dividing by two.

$$x_m = \frac{lowerboundary + uperboundary}{2} = \frac{lowerlimit + uperlimit}{2}$$

Cumulative Frequency is a distribution that shows the number of data values less than or equal to a specific value (usually an upper boundary). The values are found by adding the frequencies of the classes less than or equal to the upper class boundary of a specific class.
Cumulative Relative Frequency is running total of the relative frequencies or the cumulative frequency divided by the total frequency. Gives the percent of the values which are less than the upper class boundary.

Example 2 *the following data represent the record high temperatures in degrees Fahrenheit for 50 cities. Construct a grouped frequency distribution for the data using 7 classes.*

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	106	110
116	108	110	121	113	120	119	111	104	111
120	113	120	117	105	110	118	112	114	114

solution

First Determine the classes.
Find the highest value and lowest value: H=134 and L=100.
Find the range:$R = highestvalue - lowestvalue = H - L = 134 - 100 = 34$
Select the number of classes in this case 7 is arbitrarily chosen.
Find the class *width* by dividing the range by the number of classes.
$width = \frac{R}{numberofclasses}$
$w = \frac{34}{7} = 4.9 \approx 5$
to start the classes, each class has lower and upper limit,
for the first class the lowest values in the data can be used as the lower limit 100
to get the upper class limit add $width - 1$ to the lower limit,$100 + (5 - 1) = 104$
the second class lower limit is the first class $upperlimit + 1 = 104 + 1 = 105$
and so on until there are 7 classes
Find the ***class boundaries*** by subtracting 0.5 from each lower class limit and adding 0.5 to each upper class limit

Second Tally the data and place the results in the tally column.

Third Find the numerical frequencies from the tallies.

Fourth Find the cumulative frequencies by adding the frequencies of the classes less than or equal to the upper class boundary of a specific class.

Class limit	Class boundaries	Tally	Frequency	Cumulative frequency
100−−104	99.5−−104.5	||	2	2
105−−109	104.5−−109.5	𝍸 |||	8	10
110−−114	109.5−−114.5	𝍸 𝍸 𝍸 |||	18	28
115−−119	114.5−−119.5	𝍸 𝍸 |||	13	41
120−−124	119.5−−124.5	𝍸 ||	7	48
125−−129	124.5−−129.5	|	1	49
130−−134	129.5−−134.5	|	1	50
			Total=50	

The frequency distribution shows that the class 109.5−−114.5 contains the largest number of temperatures (18) followed by the class 114.5−−119.5 with 13 temperatures. Hence, most of the temperatures (31) fall between 109.5 and 119.5.

Rules for constructing frequency tables

- There should be between 5 and 20 classes.
- The class width preferred to be an odd number to ensure that the **class midpoint** x_m has the same places as the data
- The classes must be mutually exclusive
- The classes must be continuous
- The classes must be exhaustive
- The classes must be equal in width.

Note The classes must be equal in width to avoids a distorted view of the data. One exception occurs when a distribution has a class that is open-ended. That is, the class has no specific beginning value or no specific ending value. A frequency distribution with an open-ended class is called an **open-ended distribution**. Here an examples of distribution with open-ended classes

Class limit	Frequency
10−−20	3
21−−31	6
32−−42	4
43−−53	10
54-and-above	8

4 Ungrouped Frequency Distribution

This type of distribution use when the range of the data values is relatively small, a frequency distribution can be constructed using single data values for each class.

Example 3 *The data shown here represent the number of miles per gallon (mpg) that 30 selected four-wheel-drive sports utility vehicles obtained in city driving. Construct a frequency distribution, and analyze the distribution*

12	17	12	14	16	18
16	18	12	16	17	15
15	16	12	15	16	16
12	14	15	12	15	15
19	13	16	18	16	14

solution

First Determine the classes.
Since the range of the data set is small $(19 - 12 = 7)$,
classes consisting of a single data value can be used. They are 12, 13, 14, 15, 16, 17, 18, 19.
Second Tally the data.
Third Find the numerical frequencies from the tallies.
Fourth Find the cumulative frequencies by adding

Class limit	Class boundaries	Tally	Frequency	Cumulative frequency
12	11.5–−12.5	𝍸 |	6	6
13	12.5–−13.5	|	1	7
14	13.5–−14.5	|||	3	10
15	14.5–−15.5	𝍸 |	6	16
16	15.5–−16.5	𝍸 |||	8	24
17	16.5–−17.5	||	2	26
18	17.5–−18.5	|||	3	29
19	18.5–−19.5	|	1	30
			Total=30	

The frequency distribution shows the almost one-half (14) of the vehicles get 15 or 16 miles per gallon.

5 Summarizing Data from More Than One Variable

all the previous discussed methods for summarizing data from a single variable. Frequently, more than one variable is being studied at the same time, and we might be interested in summarizing the data on each variable separately, and also in studying relations among the variables.

5.1 Contingency table

summarizing data from two qualitative variables. The rows of the table identify the categories of one variable, and the columns identify the categories of the other variable. The entries in the table are the number of times each value of one variable occurs with each possible value of the other

Example 4 *survey was conducted on 1,272 individuals. Each individual surveyed was asked to state his or her place of residence and work performance .the result contingency table is as follow.*

Residence	*work performance*			*Total*
	week	*Moderate*	*Excellent*	
Urban	*144*	*180*	*90*	*414*
Suburban	*135*	*240*	*96*	*471*
Rural	*108*	*205*	*54*	*387*
Total	*387*	*625*	*240*	*1272*

All the different types of distributions are used in statistics and are helpful when one is organizing and presenting data. The reasons for constructing a frequency distribution are as follows:

- To organize the data in a meaningful, intelligible way.
- To enable the reader to determine the nature or shape of the distribution.
- To facilitate computational procedures for measures like the average.
- To enable the researcher to draw charts and graphs for the presentation of data.
- To enable the reader to make comparisons among different data sets.