# Cartography Lecture #7
DESCRIPTIVE STATISTICS

Asst. Prof. Dr.Haval AbdulJabbar Sadeq
Geomatics Eng. Dep., College of Engineering,
Salahaddin University-Erbil, Iraq.
2023- Spring Semester

1

Statistical Data

Prior to mapping, the data are evaluated in terms of their statistics, size, and forms. Because, the data sets used in thematic mapping come in all sizes and characteristics.

Legend
4.0 - 6.0 [56]
6.0 - 7.0 [38]
7.0 - 8.0 [19]
8.0 - 11.0 [36]
11.0 - 24.0 [27]
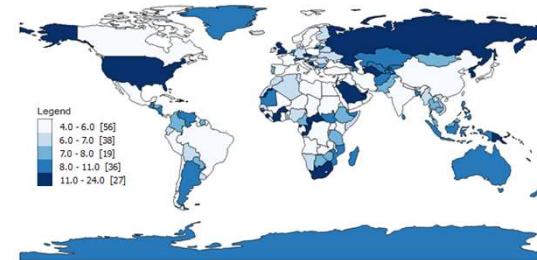
2 Cartography

Statistical Data Properties

1. Statistical data provide the cartographer with insight into a variety of issues and topics. The size of a database is directly dependent upon both the number of observations and the number of variables.
2. Data classification provides measures for simplifying large data sets to assist in analysis and display.
3. The decisions facing the cartographer involve the number of classes and the data classification scheme to be used in generalizing the data.

4. The variations in spatial patterns based on the number of classes and data classification are presented in both numeric and graphic forms.

3 Cartography

Statistical Data indices

The indices used by geographers and cartographers are. The first two are often used interchangeably, but in fact, they differ.

1. **Ratio,**

2. **Proportion,**

3. **Percent, and,**

4. **Rate**

4 Cartography

Statistical Data (Ratio)

A ratio is a good way of expressing the relationship between two data entities. It is expressed as:

$$\frac{f_a}{f_b}$$ Where: fa is the number of items (frequency) in one entity and fb is the number in a second entity.

The density of the population can be calculated from: $\frac{\textit{Number of people in a county}}{\textit{Area of county in square miles}} = \textit{People per square mile}$

Example-1: consider the ratio of the sexes enrolled in the population of a major southern university in the fall semester of 2006. There were 10,876 males and 14,248 females matriculating. The ratio is calculated as:

$$\frac{14,248}{10,876} = 1.31$$

Example-2: find that Appling County's population (POP2000) is 17,419 and the county area (AREA) is 510.3680 square miles producing this calculation:

$$\frac{17,419}{510.3680} = \frac{34.13}{1}$$

Therefore the calculated population density for this county is 34.13 people per square mile.

5 Cartography

Statistical Data (Proportion)

Proportion is the ratio of the number of items in one group (class) to the total of all items. It is written:

$$\frac{f_a}{N}$$ where f_a is the number of items (frequency) in a class and N is the total number of items or total frequency.

To determine the proportion of housing units that are owner occupied, again using Appling County's data, the calculation would be: $\frac{6,606}{7,854} = 0.8411$

Typically, proportions are multiplied by 100, yielding a **percentage.**

In this case, 0.8411*100=84.11% of the county's housing units are occupied

6 Cartography

Statistical Data (Percentage)

Percentage change is another frequently calculated variable based on a single variable for two different periods. $\frac{\text{New Value} - \text{Old Value}}{\text{Old Value}} * 100 = \text{Percent Change}$

EXAMPLE: change in population in the county

Forsyth County $\frac{98,407 - 44,083}{44,083} = \frac{54,324}{44,083}$
$= 1.2322 * 100 = 123.23\%$

Chattahoochee County $\frac{14,882 - 16,934}{16,934} = \frac{-2,052}{16,934}$
$= -0.1212 * 100 = -12.12\%$

This negative percent (-12.12%) indicates that the county lost about one-fifth of its population during the specific period and now has 2,052 fewer people living in the county

7 Cartography

Statistical Data (Rates)

Rates are similar to percentages except that the relationship is a value per some much larger value. It is determined by the relationship between an observed number compared to a potential number of occurrences for a given time or place. The result is then multiplied by a power of ten, usually relative to the denominator, to make the result meaningful

$$\frac{\textit{Number of Occurrences}}{\textit{Number of Possible Occurrences}} \times 1000^*$$

Example: The General Fertility Rate (GFR) for a given county is calculated using the formula:

(Number of Live Births in an age group / Female population ages 15–44) $\times$ 1,000

The resulting rate provides us with the number of live births to women of any age per 1,000 females ages 15–44., the data are calculated as:

$$\frac{278}{3,731} = 0.0745 * 1,000 = 74.5$$

8 Cartography

Data classification selection:

Analysis of the selection of class intervals suggested the following five requirements should be met:

1. Encompass the full range of the data.
2. Have neither overlapping values nor vacant classes.
3. Be great enough in number to avoid sacrificing the accuracy of the data.
4. Divide the data into reasonably equal groups of observations.
5. Have a logical mathematical relationship if practical.

9 Cartography

Descriptive Statistics

A common set of statistics about az variable include the data's total, maximum minimum values, data range, arithmetic mean, frequency of observations (numl of counties), and standard deviation.

The **median** is the midpoint of the data and should not be confused with the da mean.

The **variance** is : $\sigma^2 = \dfrac{\Sigma(x-\overline{X})^2}{N}$

The square root of the variance to obtain the measure of deviation, which is called standard deviation (STD) $\sigma = \sqrt{\dfrac{\Sigma(x-\overline{X})^2}{N}}$

The standard deviation is especially useful when comparing two variables. The smaller of two standard deviations indicate values occur closer to the mean than a larger standard deviation.
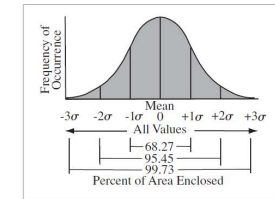
MEAN AND MEDIAN EXAMPLE

The following two hypothetical data sets can be used to depict these differences:

	Set 1	Set 2	
	8	5	
	12	8	
	14	12	
Median = 16	16	14	Median = 15
Mean = 18.3	22	16	(14 + 16 = 30/2 = 15)
	26	22	Mean = 16.6
	30	26	
		30	
	128	133	

Frequency of Occurrence

Mean
-3σ -2σ -1σ 0 +1σ +2σ +3σ
All Values
— 68.27 —
— 95.45 —
— 99.73 —
Percent of Area Enclosed

Cartography

Thank you