



الكلية : كلية طب العام

الفرع : طب الاسرة والمجتمع

المرحلة : الثالثة

أستاذ المادة : د بديعه ثامر يحيى

اسم المادة باللغة العربية : الاحصاء الحياتي

اسم المادة باللغة الانكليزية : Biostatistics

اسم المحاضرة الاولى باللغة العربية : تلخيص البيانات

اسم المحاضرة الاولى باللغة الانكليزية : Summarizing of data

Dr. Badeaa Thamer Yahyaa

Biostatistics (2023 – 2024) L1

Summarizing of data

Statistics:

**is the science of conducting studies to
collect, organize, summarize, analyze,
present interpret, and draw
conclusions from data**

Data \ sample : (X_1, \dots, X_n).

Variable: is a characteristic or condition that can change or take on different values

Numerical (quantitative) variable: if a numerical quantity is assigned to each item in the sample

- 1- **Discret:** number of equipment in the project
number of students in the class (0.5,10,15,20,25, etc.)
- 2- **Continues:** height, weight, age

Categorical (qualitative) variable: if the sample items are placed into categories

1- **Nominal:** categories have no ordering

(sex: male, female\ hair color)

2- **Ordinal:** categories have ordered (grade: A, B, C, D \
rating: high, low)

Frequency distributions for organizing and summarizing Data Definitions :

Frequency distribution (frequency Table): list data values (either individually or by groups of intervals) , along with their corresponding frequencies.

Lower class limit: The smallest numbers that can belong to each of the different classes.

Upper-class limit: The largest number that can belong to each of the different classes.

Class Boundaries: The numbers used to separate the classes without the caps created by the class limit.

Class midpoints: The values in the middle of classes. Each class midpoint can be found by adding the lower-class limit to the upper-class limit and dividing the sum by 2.

Class width: The difference between two consecutive class limits in a frequency distribution.

Procedure for construction frequency distribution

1- select numbers of class intervals usually between 5 – 20 by Sturges rule.

* Those who need more specific guidance in the matter of deciding how many class intervals to employ may use a formula given by Sturges.

This formula gives $k = 1 + 3.322 \log (n)$ where k stands for the number of class intervals and n is the number of values in the data set under consideration.

$$L_x - S_x$$

$$\text{Max} - \text{Min}$$

2- Calculate the class width: $W = \frac{\text{-----}}{K}$ OR $\frac{\text{-----}}{\text{no. of classes}}$

3- choose the value for the first lower class limit by using the minimum value.

4- Using the first lower class limit and class width.

5- List the lower-class limit in the vertical column and then determine and enter the upper-class limit.

6- Take each individual data value and put it in the appropriate class (frequency).

Example: In a certain study for diastolic blood pressure for 35 patients visit a community health center :

70	80	60	80	80
60	90	60	70	70
70	60	80	60	60
80	90	120	80	
60	100	60	90	
60	60	90	70	
80	90	100	80	
60	80	100	90	

$$K = 1 + 3.322 \log(n)$$

$$k = 1 + 3.322 \log(35)$$

$$K = 1 + 3.322 (1.544)$$

$$K = 6.1 \approx 6$$

Diastolic blood pressure	FREQUENCY
60 – 69	11
70 – 79	5
80 – 89	9
90 – 99	6
100 – 109	3
110 - 119 OR 110 – 120	1
TOTAL	35

$$w = \frac{Lx - Sx}{K} = \frac{120 - 60}{6} = 10$$

Diastolic blood pressure	FREQUENCY	Diastolic blood pressure	f	Diastolic blood pressure	F	Diastolic blood pressure	F
60 – 69	11	60-	11	60 <	11	60>	11
70 – 79	5	70-	5	70<	5	70>	5
80 – 89	9	80 -	9	80 <	9	80>	9
90 – 99	6	90-	6	90<	6	90>	6
100 – 109	3	100-	3	100<	3	100>	3
110 - 119 OR 110 – 120	1	110-	1	110<	1	110>	1
TOTAL	35	total	35	total	35	Total	35

Diastolic blood pressure	F	RF %	Cf	CRF
60 -	11	$11 \div 35 \times 100 = 31.4$	11	31.4
70-	5	14.3	16	45.7
80-	9	25.7	25	71.4
90-	6	17.1	31	88.5
100-	3	8.6	34	97.1
110 -	1	2.9	35	100
Total	35	100		

- **Relative Frequency** It may be useful at times to know the proportion, rather than the number, of values falling within a particular class interval. We obtain this information by dividing the number of values in the particular class interval by the total number of values.
- **Cumulative Frequency** is the sum of class and all classes below it in frequency distribution. All it means is yours adding up a value and all of the values that come before it.
- **Cumulative Relative frequency** is a statistical calculation figured by adding together previously tabulated relative frequencies that makes a running total a long frequency Table.