Organizing Data Using Tables and Graphs

Class Notes by A.Sandanasamy, BHC.

ORGANIZING DATA

Raw, Unorganized Leadership Scores of 35 Managers:

54	43	48	50	52	44	49
44	46	51	42	50	46	50
51	55	57	48	53	51	48
46	52	50	45	55	49	53
48	50	49	51	48	43	52

SIMPLE FREQUENCY DISTRIBUTIONS

A **simple frequency distribution simply lists the frequencies with which each raw score occurs.**

X – raw score
f - frequency
N - total number of scores (Population)
n – total number of scores (Sample)

X	Tally	f
57	|	1
56		0
55	||	2
54	|	1
53	||	2
52	|||	3
51	||||	4
50	𝍸	5
49	|||	3
48	𝍸	5
47		0
46	|||	3
45	|	1
44	||	2
43	||	2
42	|	1
		$N = 35$

To Create a Simple Frequency Distribution:

1. Create labels for three columns, as follows: X, Tally, and f.
2. Locate the highest and lowest scores in the unorganized list of scores.
3. Beginning with the highest score at the top, list the score values in descending order in the X column of your frequency distribution. Stop at the lowest obtained score.
4. Underline the first score in your unorganized list and place a tally mark for that score in the Tally column of your frequency distribution.
5. Count the number of tally marks for each score and record this number in the f column.
6. Total the scores in the f column. The sum should be equal to the total number of scores (N).

RELATIVE FREQUENCY DISTRIBUTIONS

The frequency refers to how many times a particular score occurs, whereas in a relative frequency distribution, the frequency refers to the proportion of time that the score occurs

Rel. $f = f / N$

X	f	*Rel. f*
10	4	.15
9	2	.08
8	6	.23
7	4	.15
6	5	.19
5	3	.12
4	0	.00
3	2	.08
	$N = 26$	1.00

CUMULATIVE FREQUENCY DISTRIBUTIONS

Cumulative frequency distribution indicates the frequency of scores that fall at or below a particular score value.
This type of table is useful if you want to know how many people scored below a certain value on a test.

To Create a Cumulative Frequency Distribution:

1. **Begin with a simple frequency distribution table.**
2. **Add a cumulative frequency (*cf) column to the table.***
3. ***Work from the bottom up in the f and cf columns. Take the bottom score in the f column and enter that number in the corresponding space in the cf column. Then take that num-ber (in the cf column) and add it to the next number up in the f column and record the total in the next space up in the cf column. Take the last number entered in the cf column and add it to the next number up in the f column. Repeat this process until the cf column is complete. The number at the top of the cf column should be equal to N.***

Cumulative Frequency Distribution for Scores on the Statistics Quiz:

X	f	cf	
10	4	26	(the cf at the top is equal to N)
9	2	22	
8	6	20	
7	4	14	
6	5	10	
5	3	5	(cf of 2 below + f of 3 = 5)
4	0	2	(because no students scored 4, the cf is still 2)
3	2	2	
	$N = 26$		

PERCENTILE RANK

Percentile rank (P.R.) gives a bit more information than just frequency by indicating the percentage of scores that fall at or below a given score in a distribution.

$$P.R. = \frac{cf}{N} \times 100$$

Percentile Rank for Scores on the Statistics Quiz:

X	f	cf	$P.R.$
10	4	26	100.00
9	2	22	84.62
8	6	20	76.92
7	4	14	53.85
6	5	10	38.46
5	3	5	19.23
4	0	2	7.69
3	2	2	7.69
	$N = 26$		

COMBINING THE TABLES

Complete the columns that follow for the raw scores below:

29	33	32	34
32	32	31	32
33	32	32	33
32	31	34	32
31	33	33	31

X	*Tally*	*f*	*Rel. f*	*cf*	*P.R.*
____	____	____	____	____	____
____	____	____	____	____	____
____	____	____	____	____	____
____	____	____	____	____	____
____	____	____	____	____	____
____	____	____	____	____	____

GROUPED FREQUENCY DISTRIBUTIONS

Grouped frequency distributions, on the other hand, combine scores into groups, referred to as class intervals, thus condensing the data and making overall trends more apparent.

Class interval	*f*
81–83	1
78–80	3
75–77	6
72–74	11
69–71	7
66–68	9
63–65	21
60–62	15
57–59	13
54–56	4
51–53	3
48–50	5
45–47	2
	$N = 100$

TERMS

Class intervals refer to groups of scores, such as 6–7, 21–23, or 55–59.

Interval size, symbolized by i, refers to the number of scores in the class interval. For example, i = 3 for the class interval of 21–23 because there are three scores in the interval.

The range, symbolized by R, refers to the amount of spread in a distribution of scores. It is determined by subtracting the lower limit (LL) of the lowest score from the upper limit (UL) of the highest score.

For example, if the lowest score in a distribution is 53 and the highest score is 81, then the range for that distribution would be 29 (*R = 81.5 − 52.5).*

Ground Rules

We will use between 10 and 20 class intervals.

We will use interval sizes (*i) of 2, 3, or 5.*

*Begin the class interval column at the bottom of the distribution with a multiple of i , (*For instance, if *i = 3 in a distribution where the lowest raw score is 31, the first class interval should be 30–32 because 30 is a multiple of 3 and 31 is not)*

The largest scores go at the top of the distribution.

Bowling Scores of a High School Gym Class:

82	128	110	140	127
109	92	119	142	111
126	85	124	138	92
83	114	**146**	112	86
98	132	128	95	120
122	115	92	116	119
79	113	81	112	115

Step 1: Find the range. The high score is 146 and the low score is 79.
Thus, $R = 146.5 - 78.5 = 68$.

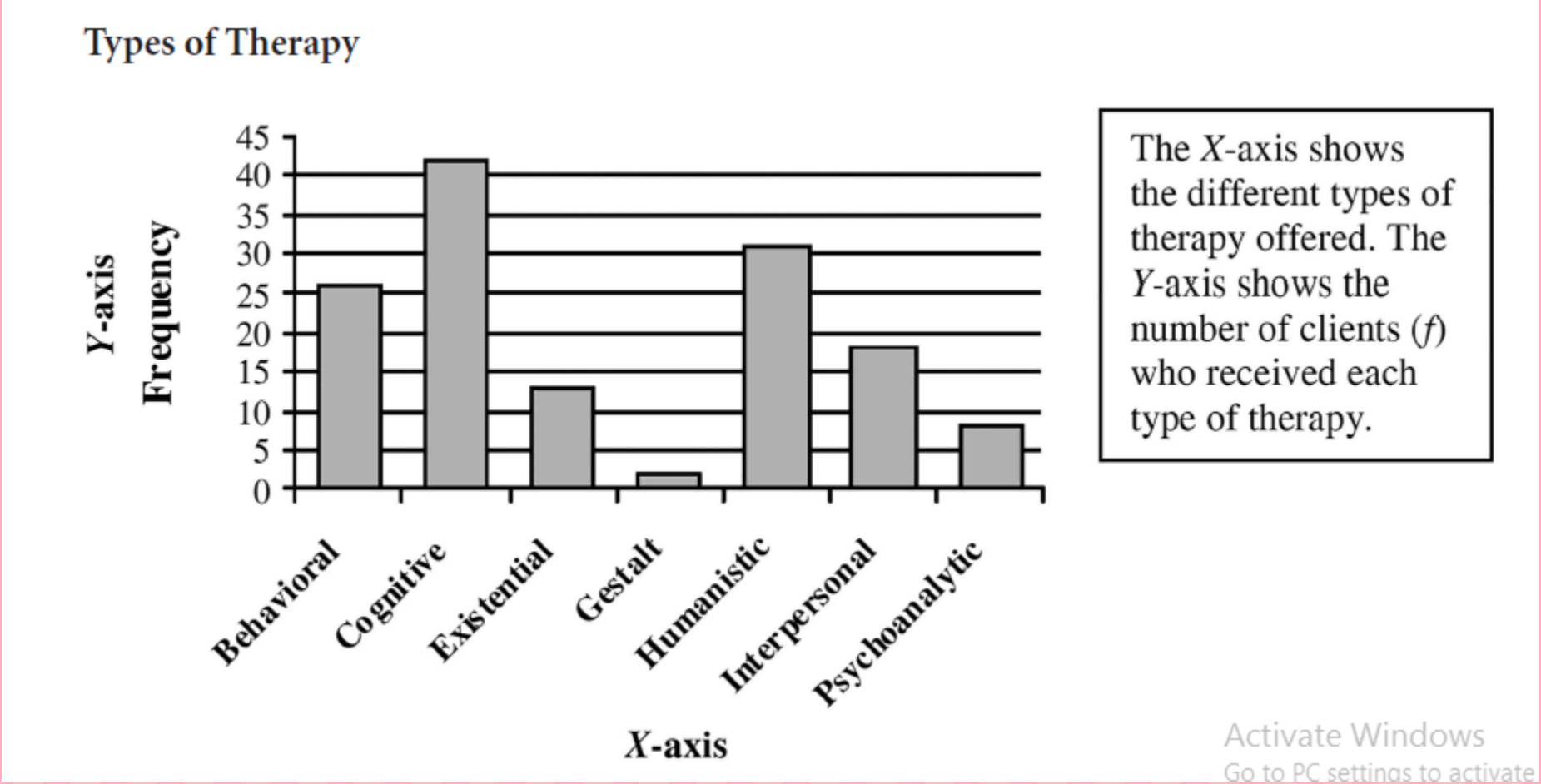
Step 2: Determine the size of the class interval (i) by trial and error.
$68 \div 2 = 34$ (too many)
$68 \div 3 = 22.67$ (too many)
$68 \div 5 = 13.60$ (just right, $i = 5$)

Step 3: Create labels.

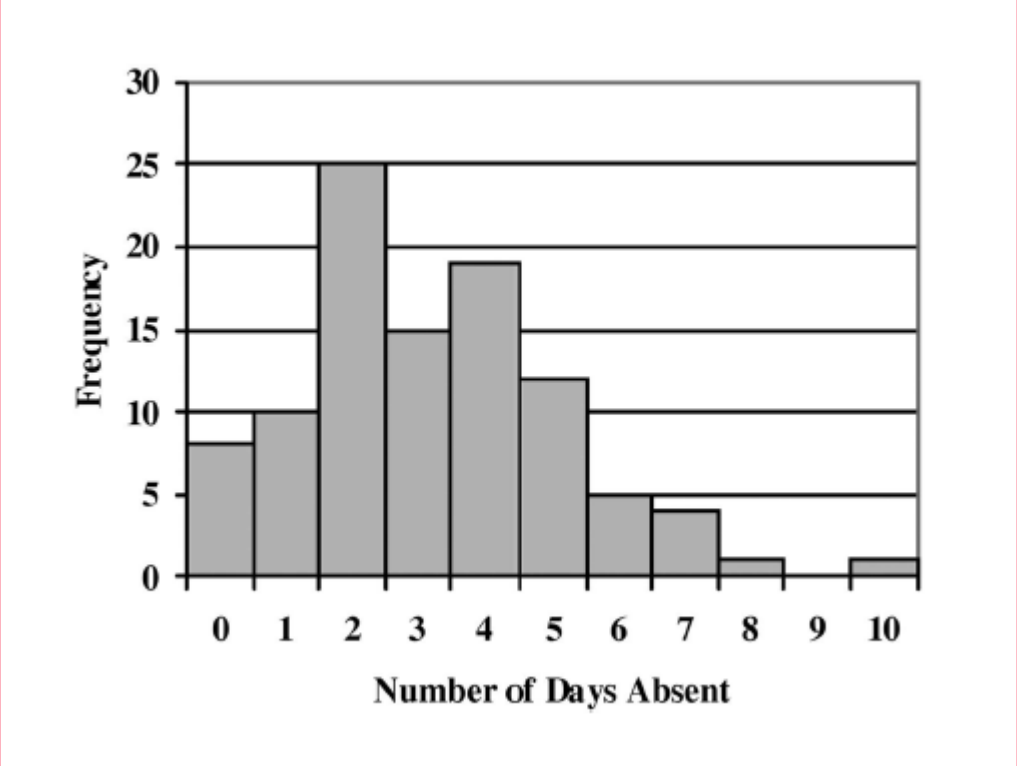
Class interval	*Tally*	*f*
145–149	|	1
140–144	||	2
135–139	|	1
130–134	|	1
125–129	||||	4
120–124	|||	3
115–119	||||	5
110–114	|||| |	6
105–109	|	1
100–104		0
95–99	||	2
90–94	|||	3
85–89	||	2
80–84	|||	3
75–79	|	1
	$N = 35$	

Steps 4 & 5: Beginning at the bottom, create your class intervals.

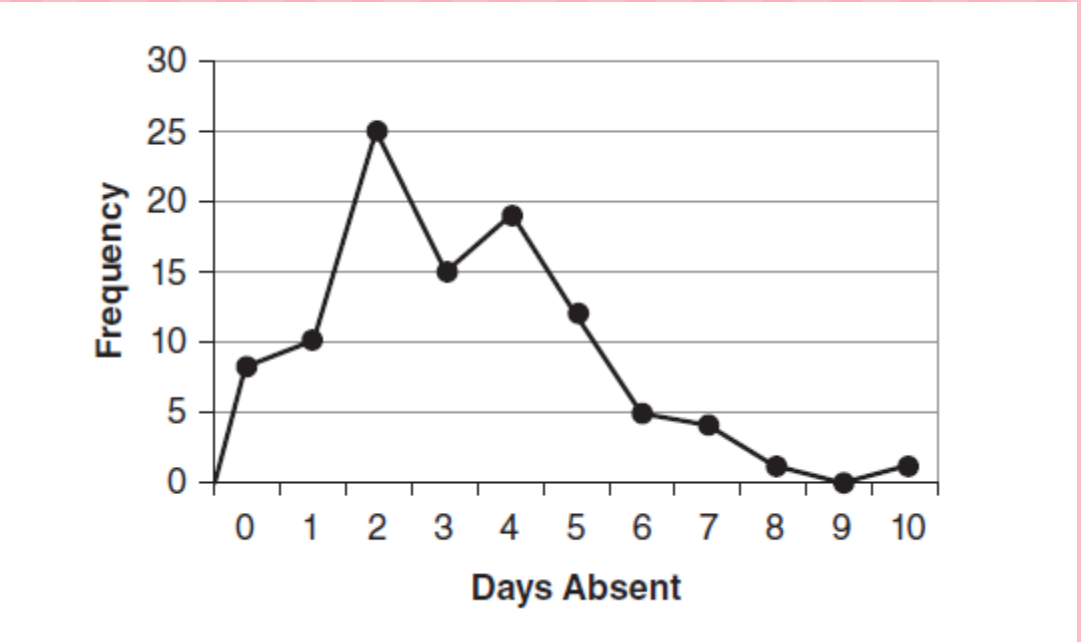
GRAPHIC REPRESENTATION OF DISTRIBUTIONS



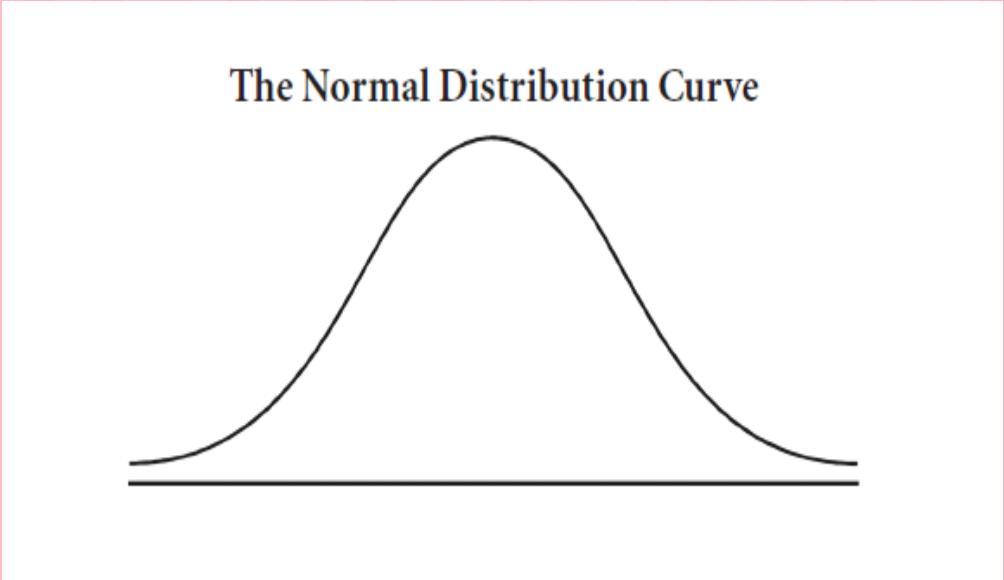
Histograms are used for quantitative variables that differ in amount. These would include data measured on interval or ratio scales. Histograms are similar to bar graphs except that the bars touch each other to reflect the continuous nature of the measured variable



Frequency polygons are also used for quantitative data; dots, corresponding to the appropriate frequency, are used instead of bars. The dots are then connected by a straight line.



TYPES AND SHAPES OF FREQUENCY POLYGONS



Some important characteristics of the bell curve are as follows:

- It is symmetrical, with the left and right sides being mirror images of each other.
- It is based on an infinite number of cases. Therefore, the tails are ***asymptotic***, meaning that they never touch the baseline (i.e., a theoretical normal distribution can never have a frequency of zero).
- The most frequently occurring scores are in the middle.
- The least frequently occurring scores are farthest from the middle.

Skewed distributions, on the other hand, are not symmetrical