LECTURE 3
DESCRIPTIVE STATISTICS: FREQUENCY DISTRIBUTIONS

PSY2002

Hye Won Suk

DESCRIPTIVE STATISTICS

- Descriptive statistics (기술통계) indicate statistical procedures used to summarize, organize, and simplify data.

- Usually, a data set consists of a bunch of scores or measurements. Researchers need to condense a set of scores or measurement into a more manageable and comprehensible form so that any patterns and characteristics in the data can be seen easily and communicated to others. Descriptive statistics serve for this purpose.

DESCRIPTIVE STATISTICS

- In this course we will learn the following three types of descriptive statistics.

  - Frequency distributions (빈도 분포)
  - Central tendency measures (중심경향치)
  - Variability measures (변산성 측정치)

FREQUENCY DISTRIBUTION

- One of the most common descriptive statistics that we can use is to summarize scores in a frequency distribution.

- A <u>FREQUENCY DISTRIBUTION</u> (빈도 분포) is an organized tabulation of the number of individuals in each category on the scale of measurement.

- <u>FREQUENCY</u> (빈도) is the number of times a particular score or category occurs, i.e., count.

- A frequency distribution presents a picture of how the individual scores are distributed on the measurement scale.

- A frequency distribution can be given as a frequency table (빈도표) or a frequency graph (빈도 그래프).

- In either case, a frequency distribution provides the following two kinds of information.

  - The set of categories (or numbers) that make up the original measurement scale.

  - The number of individuals (or cases) in each category (frequency or count).

HOW TO MAKE A FREQUENCY TABLE?

- Let's consider the following two situations.

    - How to make a frequency table for the responses on a variable when the variable is measured on a <u>nominal or ordinal scale (i.e., qualitative data)</u>.

    - How to make a frequency table for the responses on a variable when the variable is measured on <u>an interval or ratio scale (i.e., quantitative data)</u>.

NOMINAL/ORDINAL

- Case 1: For a nominal or ordinal scale variable

  - STEP 1: Find unique categories.
  - STEP 2: List the unique categories in the first column of the table.
  - STEP 3: Put the frequency associated with each category in the second column.
  - STEP 4: Add other columns such as proportion and % frequency (This step is optional.)

NOMINAL/ORDINAL

- Example

    - Below are the letter grades of the 15 students who took PSY6001 last year.
    - Make a frequency table for the data.

B+, B-, A+, A, A, C, B+, A-, B, B, C+, A, A+, B, B-

NOMINAL/ORDINAL

- Basically, a frequency table has two columns as shown below:

X	f
A+	2
A	3
A-	1
B+	2
B	3
B-	2
C+	1
C	1

$$n = \sum f = 2 + 3 + 1 + 2 + 3 + 2 + 1 + 1 = 15$$

NOMINAL/ORDINAL

- If necessary, we can add other columns in the frequency table.

X	f	$p = f/n$	$\%f = p(100)$
A+	2	2/15 = 0.13	13%
A	3	3/15 = 0.20	20%
A-	1	1/15 = 0.07	7%
B+	2	2/15 = 0.13	13%
B	3	3/15 = 0.20	20%
B-	2	2/15 = 0.13	13%
C+	1	1/15 = 0.07	7%
C	1	1/15 = 0.07	7%

NOMINAL/ORDINAL

- We can also add cumulative frequency as follows:

X	f	cumulative f
A+	2	15 (=13+2)
A	3	13 (=10+3)
A-	1	10 (=9+1)
B+	2	9 (=7+2)
B	3	7 (=4+3)
B-	2	4 (=2+2)
C+	1	2 (=1+1)
C	1	1

$$n = \sum f = 2+3+1+2+3+2+1+1 = 15$$

POP QUIZ

- Let's assume that we measured the universities the students taking a class belong to.

X	f
Sogang	25
Yonsei	10
Ewha Womans	5
Catholic	5

- Does it make sense to obtain cumulative frequencies for this data set? If so, how?

INTERVAL/RATIO

- Case 2: For an interval or ratio scale variable

  - Example 1: Number of nights drinking during a typical week by college students

    > 1, 3, 2, 2, 4, 1, 1, 5, 5, 3, 4, 3, 7, 3, 3, 2, 2,
    >
    > 0, 4, 6, 3, 1, 1, 7, 3, 6, 3, 2, 6, 6, 0, 5, 4, 2

  - There are 8 unique numbers.
  - Make a frequency table as we did for an ordinal variable.

INTERVAL/RATIO

- Case 2: For an interval or ratio scale variable

  - Example 2: An instructor observed the set of 20 exam scores shown below.

    76, 85, 82, 91, 75, 84, 61, 77, 92, 86,
    58, 72, 70, 89, 60, 78, 60, 93, 73, 57

  - There are 19 unique numbers! When a set of data covers a wide range of values, it is unreasonable to list all of the scores in a frequency distribution table.

  ➔ grouped frequency distribution table

INTERVAL/RATIO

- To create a grouped frequency table, you should determine class intervals.

  - Find the minimum and maximum scores.

    | 76, 85, 82, 91, 75, 84, 61, 77, 92, 86, |
    | 58, 72, 70, 89, 60, 78, 60, 93, 73, 57 |

    - Minimum: 57
    - Maximum: 93

- Calculate the range of the scores.

  Range = Maximum – Minimum = 93 – 57 = 36

- Determine the interval width.
  - Interval width = 2
    - 36/2 = 18
    - About 18 intervals will be used.
  - Interval width = 5
    - 36/5 = 7.2
    - About 7 or 8 intervals will be used.
  - Interval width = 10
    - 36/10 = 3.6
    - About 3 or 4 intervals will be used.

- In general, the interval width yielding around 10 intervals is used for grouped frequency table. However, there is no single right answer for the optimal interval width.
- Let's assume that we decided to use the width of 5.

- Each interval is defined by its bottom score and its top score.

- The bottom score in each interval should be a multiple of the width.

  - If you are using a width of 5 points, the intervals should start with 0, 5, 10, 15, and so on.

  - If you are using a width of 2 points, the intervals should start with 0, 2, 4, 6, and so on.

  - In our example, the minimum score is 57. Therefore, the lowest interval should start with 55 (the highest multiple of 5 that is lower than 57).

INTERVAL/RATIO

- List the class intervals in the first column of the table (from highest to lowest).

X

- First, list the <u>lower bound (bottom score)</u> of each interval.
- Then list the <u>upper bound (top score)</u> of each interval. The upper bound for each interval equals the lower bound for next interval – minimum measurement unit.

INTERVAL/RATIO

- Put the frequency associated with each interval in the second column.

X	f
90-94	3
85-89	3
80-84	2
75-79	4
70-74	3
65-69	0
60-64	3
55-59	2

INTERVAL/RATIO

- If necessary, we can add other columns in the frequency table.

X	f	$p = f/n$	$\%f = p(100)$
90-94	3	3/20 = 0.15	15%
85-89	3	3/20 = 0.15	15%
80-84	2	2/20 = 0.10	10%
75-79	4	4/20 = 0.20	20%
70-74	3	3/20 = 0.15	15%
65-69	0	0/20 = 0.00	0%
60-64	3	3/20 = 0.15	15%
55-59	2	2/20 = 0.10	10%

INTERVAL/RATIO

- If necessary, we can add cumulative frequency.

X	f	*cumulative f*
90-94	3	20 (=17+3)
85-89	3	17 (=14+3)
80-84	2	14 (=12+2)
75-79	4	12 (=8+4)
70-74	3	8 (=5+3)
65-69	0	5 (=5+0)
60-64	3	5 (=2+3)
55-59	2	2

HOW TO MAKE A FREQUENCY GRAPH?

- Now, let's think about how to make a graph for a frequency distribution.

- Again, consider the following two situations.

  - How to make a frequency graph for the responses on a variable when the variable is measured on a <u>nominal or ordinal scale</u>.

  - How to make a frequency graph for the responses on a variable when the variable is measured on an <u>interval or ratio scale</u>.
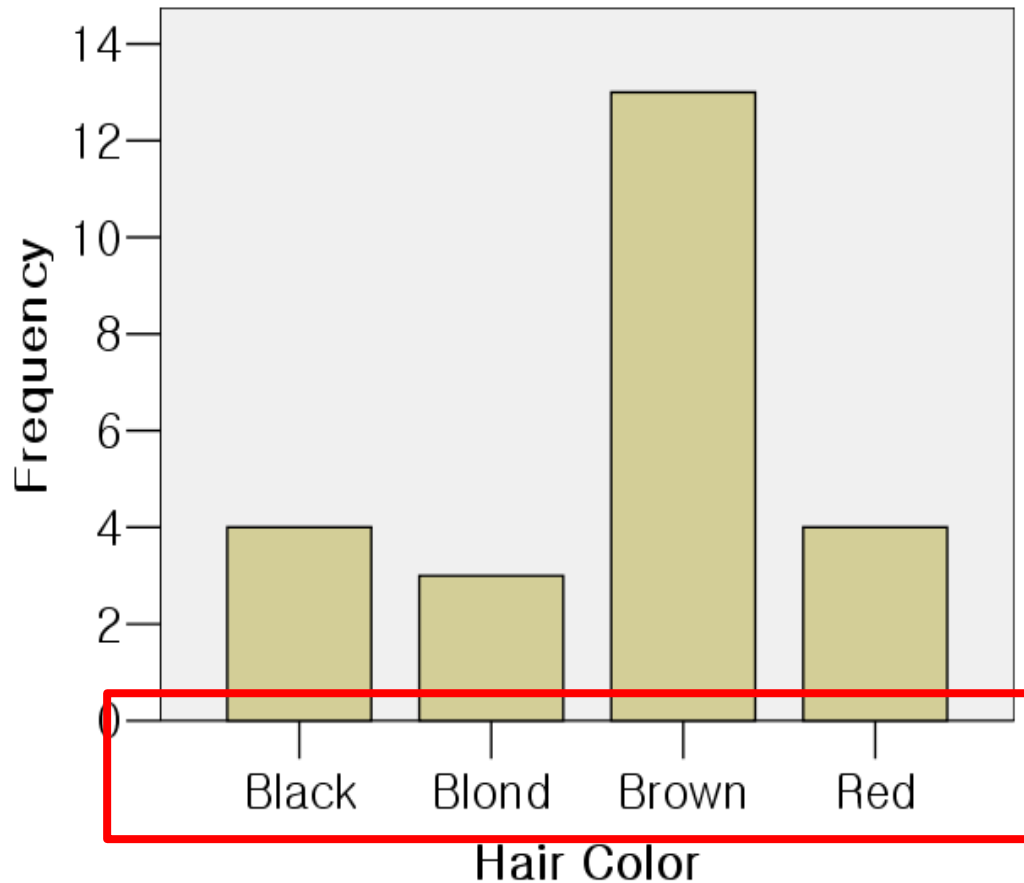
NOMINAL/ORDINAL

- Example: Hair colors

X	f
Black	4
Blond	3
Brown	13
Red	4

NOMINAL/ORDINAL

- Bar graph

The bars are not "touching." This conveys the fact that our variable consists of a set of categories that are not continuous.

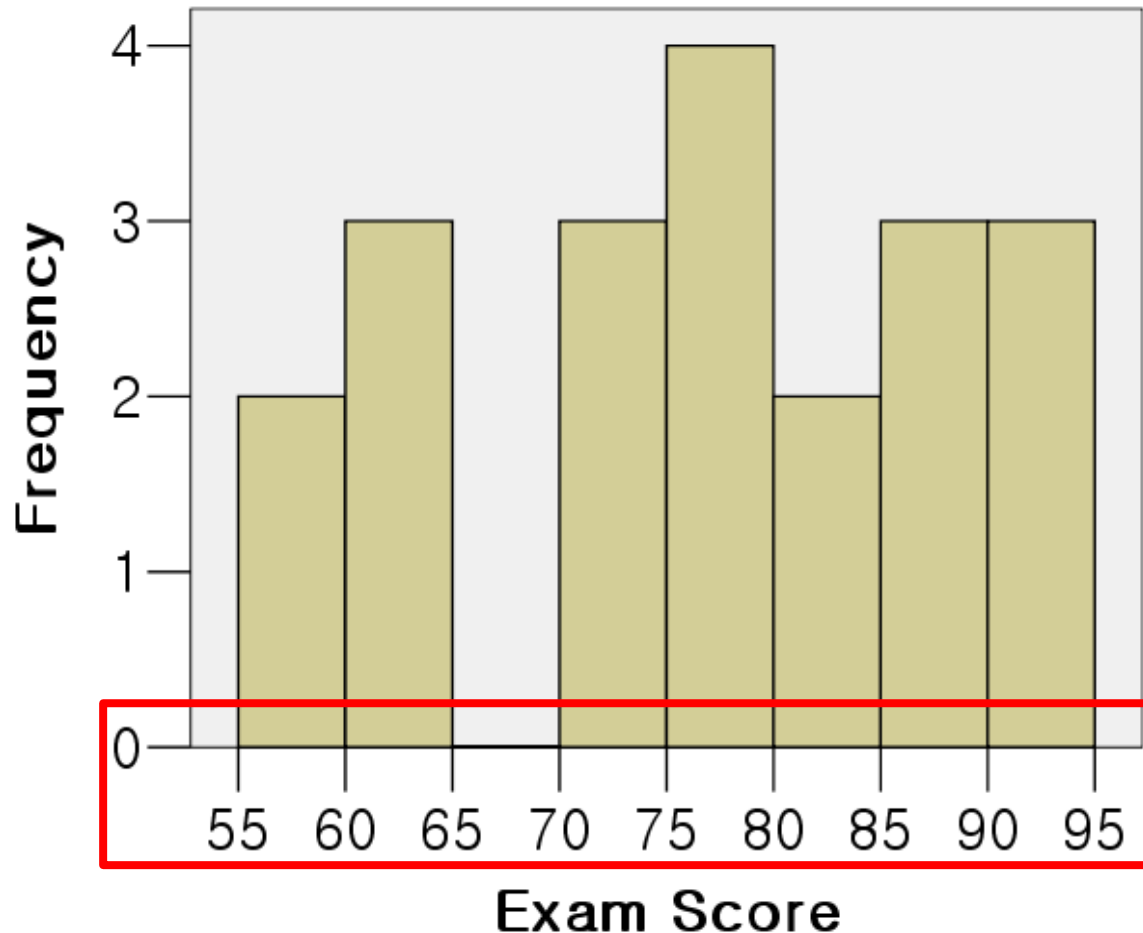The name of each category is placed at the middle of each bar.

INTERVAL/RATIO

- Example: Exam scores

X	f
90-94	3
85-89	3
80-84	2
75-79	4
70-74	3
65-69	0
60-64	3
55-59	2

INTERVAL/RATIO

- Histogram

There is no space between two adjacent bars.

The lower bound of each interval is placed at the left end of each bar.

INTERVAL/RATIO

- Frequency polygon
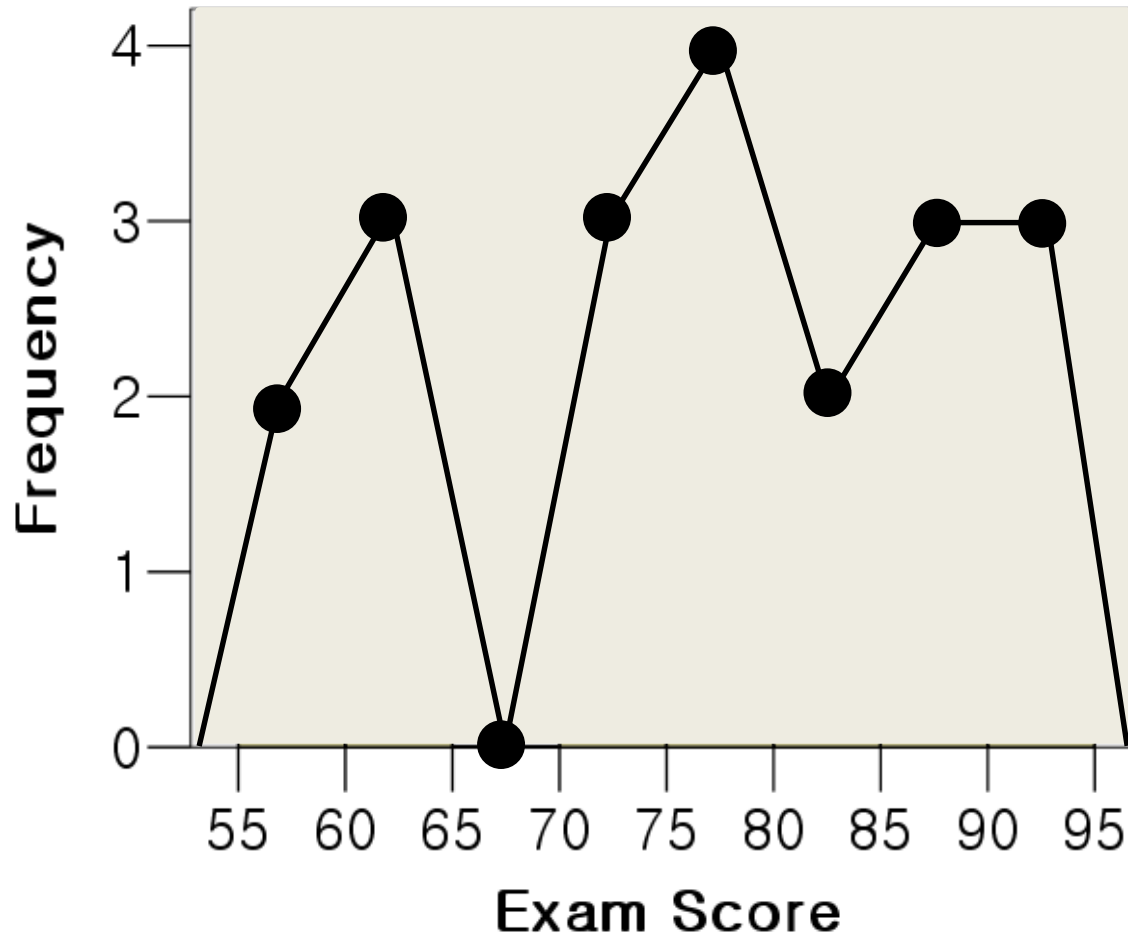 - We can make a frequency polygon from the histogram.
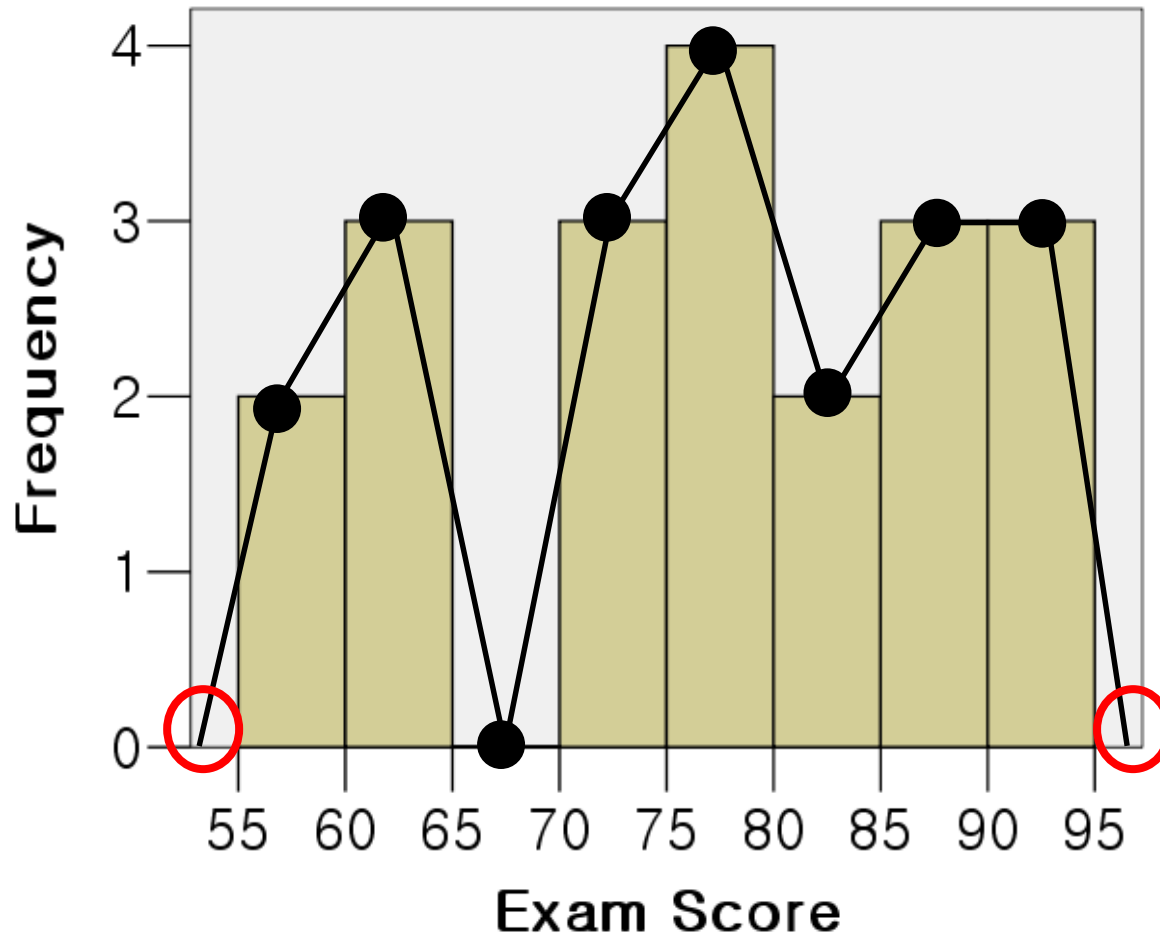
Place the dots at the midpoint of the end of each bar and connect them.

No dots are placed beyond the lowest and highest intervals.

INTERVAL/RATIO

- Frequency polygon
  - We can make a frequency polygon from the histogram.

Then delete the histogram to obtain a frequency polygon.

SUMMARY

- Frequency distribution is one of the most common descriptive statistics that shows how the scores are distributed on the measurement scale.

- Frequency table or graph can be constructed.

	Frequency table	Frequency graph
Nominal/Ordinal	• Regular frequency table	• Bar graph
Interval/Ratio	• When the number of response categories is not large: Regular frequency table • When the number of response categories is large: Grouped frequency table	• Histogram • Frequency polygon