ÇAĞ UNIVERSITY
FACULTY OF ECONOMICS & ADMINISTRATIVE SCIENCES

STATISTICS FOR BUSINESS LECTURE NOTES 4

Dr. Elma Satrovic|
elmasa1991@gmail.com

Data for a sample of 55 members of the Baseball Hall of Fame in Cooperstown, New York, are shown here. Each observation indicates the primary position played by the Hall of Famers: pitcher (P), catcher (H), 1st base (1), 2nd base (2), 3rd base (3), shortstop (S), left field (L), center field (C), and right field (R).

L	P	C	H	2	P	R	1	S	S	1	L	P	R	P
P	P	P	R	C	S	L	R	P	C	C	P	P	R	P
2	3	P	H	L	P	1	C	P	P	P	S	1	L	R
R	1	2	H	S	3	H	2	L	P					

a. Are these data qualitative or quantitative?

b. Use Excel to develop frequency distribution (COUNTIF function), relative frequency, percentage frequency, pie chart (for relative frequency) and bar chart (for frequency).

c. What position provides the most Hall of Famers?

The test scores for 42 students are summarized below.

SCORES		
12	87	50
23	75	55
23	45	65
45	55	45
56	56	44
67	50	35
88	23	40
80	65	50
90	75	60
34	80	70
37	85	66
65	77	32
45	70	34
67	52	45

a. Construct a frequency distribution. (class width is 10)

b. Construct a percent frequency distribution.

c. Construct a relative frequency distribution.

Summarizing **Quantitative** Data

Graphical Display and Tabular Summary

1. Dot Plot

☐ One of the simplest graphical summaries of data is a **dot plot.** A horizontal axis shows the range for the data. Each data value is represented by a dot placed above the axis.

☐ **Table 4.1** summarizes data on time in days required to complete year-end audits for a sample of 20 clients and frequency:
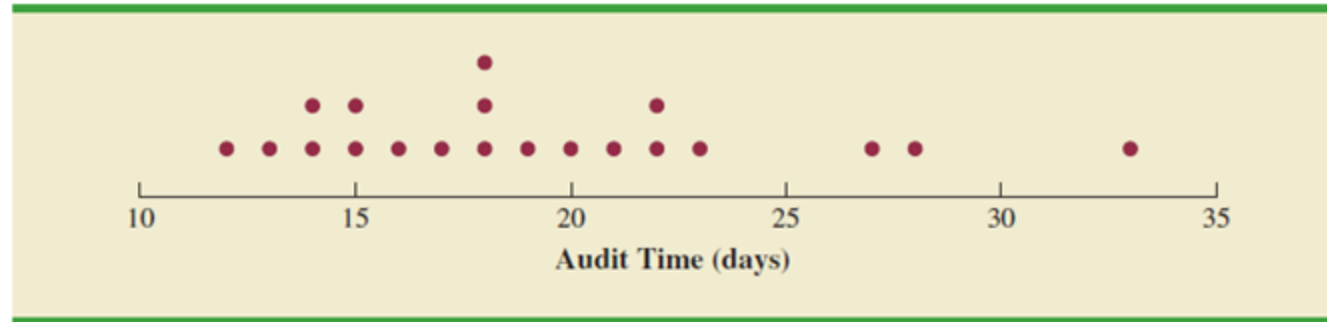
Table 4.1

YEAR-END AUDIT TIMES (IN DAYS)

12	14	19	18
15	15	18	17
20	27	22	23
22	21	33	28
14	18	16	13

Audit time	Frequency	Audit time	Frequency
12	1	23	1
13	1	24	0
14	2	25	0
15	2	26	0
16	1	27	1
17	1	28	1
18	3	29	0
19	1	30	0
20	1	31	0
21	1	32	0
22	2	33	1

□ Figure 4.1 is the dot plot for the data given in Table 4.1.

Figure 4.1 DOT PLOT FOR THE AUDIT TIME DATA

□ The three dots located above 18 on the horizontal axis indicate that an audit time of 18 days occurred three times.

□ Dot plots show the details of the data are useful for comparing the distribution of the data for two or more variables.

2. Histogram

□ A common graphical presentation of quantitative data is a **histogram,** which can be prepared for data summarized in a **frequency, relative frequency or percentage frequency** mentioned in previous lecture.

□ In order to present histogram, data on people's age from previous lecture are used:

Table 3 1 Frequency distribution for the people's age

Age (years)	Frequency
9-18	7
19-28	20
29-38	16
39-48	21
49-58	13
59-68	9
69-78	6
79-88	5
89-98	2
99-108	1
Total	100

Table 3. 2 Relative and Percentage frequency distributions

Age (years)	Relative Frequency	Percentage Frequency
9-18	0.07	7
19-28	0.2	20
29-38	0.16	16
39-48	0.21	21
49-58	0.13	13
59-68	0.09	9
69-78	0.06	6
79-88	0.05	5
89-98	0.02	2
99-108	0.01	1

☐ The variable of interest is placed on the horizontal axis, and the frequency, relative frequency or percentage frequency on the vertical axis.

☐ The frequency, relative frequency and percentage frequency of each class is shown by drawing a rectangle whose base is determined by the class limits on the horizontal axis and whose height is the corresponding frequency, relative frequency or percentage frequency.

☐ Figure 4.2 is a histogram for the frequency data of peoples' age (in Table 3.2):

Figure 4 2 Histogram for people's age
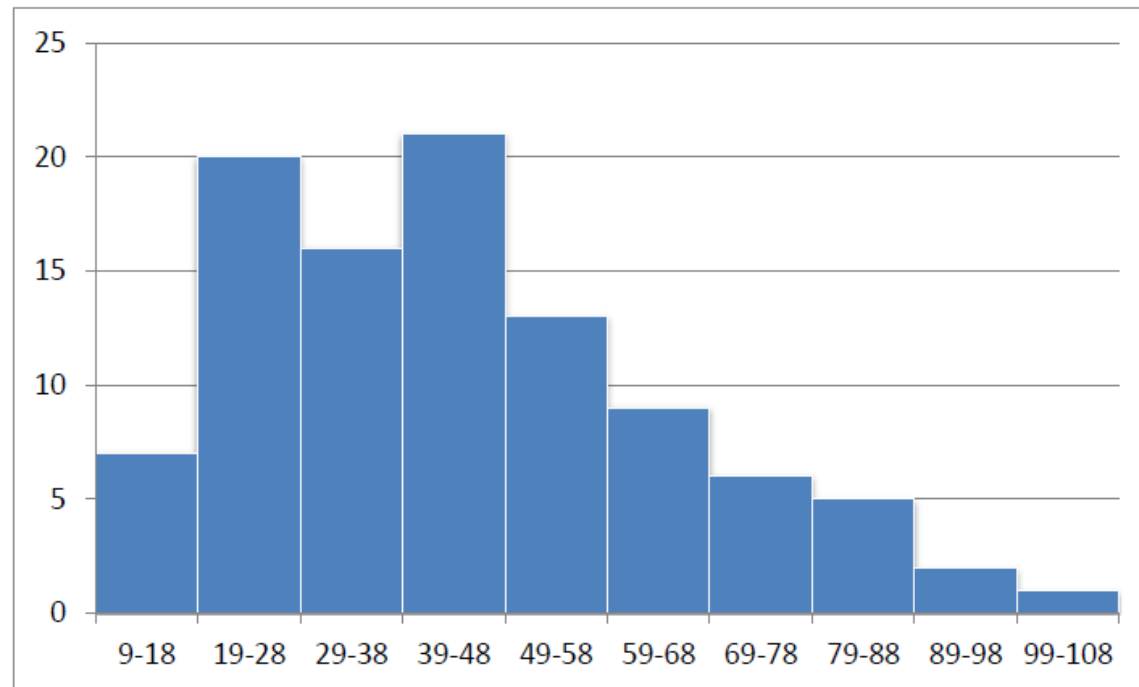
□ The adjacent rectangles of a histogram touch one another. This is usual convention for a histogram, unlike a bar chart.

□ One of the most important uses of a histogram is to provide information about the shape, or form, of a distribution.

□ Figure 4.3 contains four histograms constructed from relative frequency distributions.
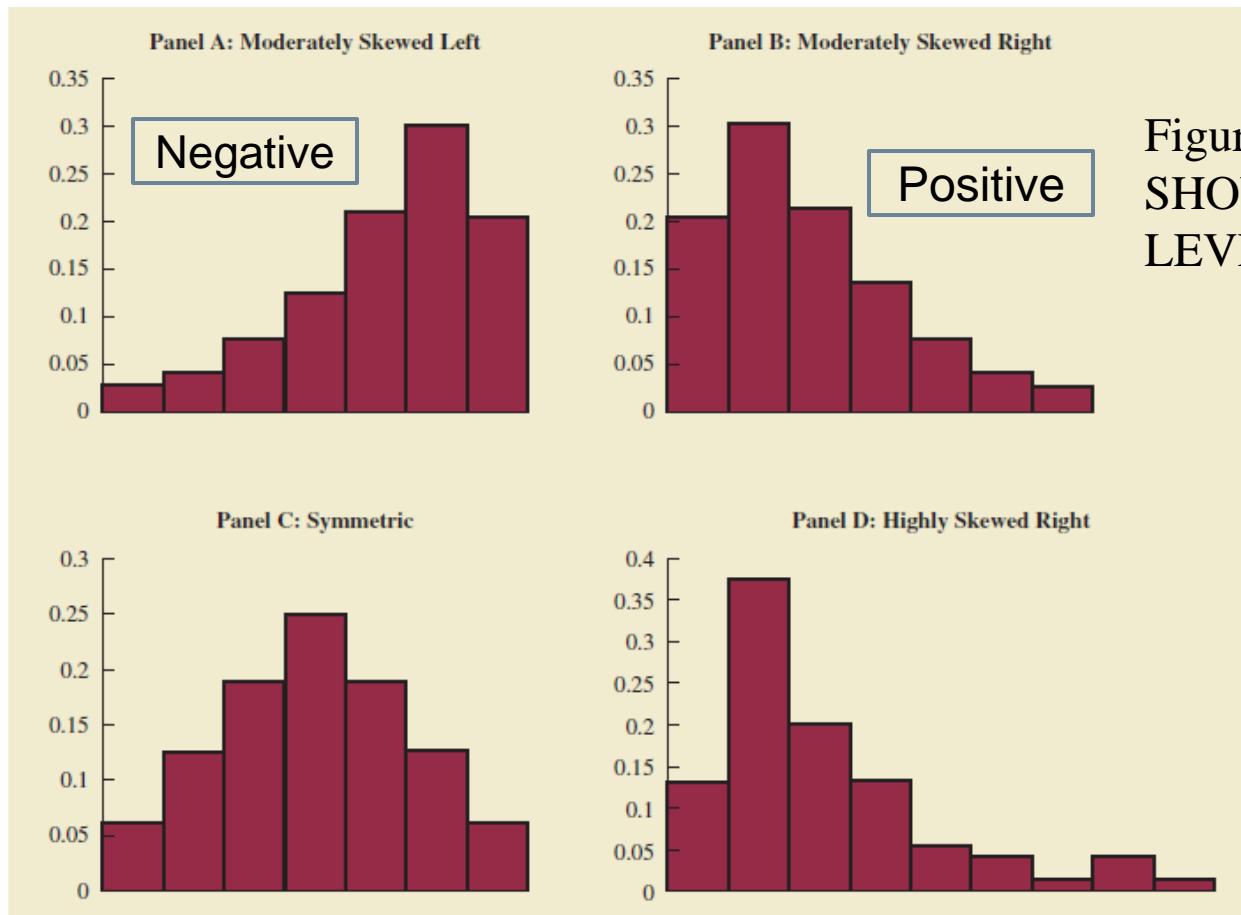

Figure 4.3 HISTOGRAMS SHOWING DIFFERING LEVELS OF SKEWNESS

- **Panel A** shows the histogram for a set of data moderately skewed to the left. A histogram is said to be skewed to the left if its tail extends farther to the left. This histogram is typical for exam scores, with no scores above 100%, most of the scores above 70%, and only a few really low scores.

- **Panel B** shows the histogram for a set of data moderately skewed to the right. A histogram is said to be skewed to the right if its tail extends farther to the right. An example of this type of histogram would be for data such as housing prices; a few expensive houses create the skewness in the right tail.

- **Panel C** shows a symmetric histogram. In a symmetric histogram, the left tail mirrors the shape of the right tail. Histograms for data found in applications are never perfectly symmetric, but the histogram for many applications may be roughly symmetric. Data for heights and weights of people lead to histograms that are roughly symmetric.

- **Panel D** shows a histogram highly skewed to the right. This histogram was constructed from data on the amount of customer purchases over one day at a women's apparel store.

- Data from applications in business and economics often lead to histograms that are skewed to the right. For instance, data on housing prices, salaries, purchase amounts, and so on often result in histograms skewed to the right.

3. Cumulative Distributions

- A variation of the frequency distribution that provides another tabular summary of quantitative data is the **cumulative frequency distribution.**

- The cumulative frequency distribution uses the number of classes, class widths and class limits developed for the frequency distribution.

- However, rather than showing the frequency of each class, the cumulative frequency distribution shows the number of data items with values less than or equal to the upper class limit of each class.

- A **cumulative relative frequency distribution** shows the proportion of data items and a **cumulative percentage frequency distribution** shows the percentage of data items with values less than or equal to the upper limit of each class.

In order to present these distributions (Table 4.2) data on people's age from previous lecture are used.

Table 4 2 Cumulative frequency, cumulative relative frequency and cumulative percentage frequency distributions for the people's age data.

Frequency Distribution		People's age	Cumulative Frequency	Cumulative Relative Frequency	Cumulative Percentage Frequency
Age (years)	Frequency				
9-18	7	Less than or equal to 18 7		0.07	7%
19-28	20	Less than or equal to 28 27		0.27	27%
29-38	16	Less than or equal to 38 43		0.43	43%
39-48	21	Less than or equal to 48 64		0.64	64%
49-58	13	Less than or equal to 58 77		0.77	77%
59-68	9	Less than or equal to 68 86		0.86	86%
69-78	6	Less than or equal to 78 92		0.92	92%
79-88	5	Less than or equal to 88 97		0.97	97%
89-98	2	Less than or equal to 98 99		0.99	99%
99-108	1	Less than or equal to 108 100		1.00	100%

Cumulative frequency interpretation

☐ Consider the class with the description 'less than or equal to 38'. The cumulative frequency for this class is simply the sum of the frequencies for all classes with data values less than or equal to 38. For the frequency distribution in Table 4.2, the sum of the frequencies for classes 9-18, 19-28 and 29-38 indicates that 7+20+16=43 data values are less than or equal to 38 meaning that 43 people are younger than or 38 years old.

Cumulative relative frequency interpretation

☐ The cumulative relative frequency distribution can be computed either by summing the relative frequencies in the relative frequency distribution or by dividing the cumulative frequencies by the total number of items.

☐ Using the latter approach, we found the cumulative relative frequencies in column 5 of Table 4.2 by dividing the cumulative frequencies in column 4 by the total number of items ($n=100$).

☐ Results show that **the proportion of people** younger than or 28 years old is 0.27.

Cumulative percent frequency interpretation

☐ The cumulative percent frequencies were again computed by multiplying the relative frequencies by 100.

☐ Results show that 77% of persons is younger that or 58 years old.

4. Ogive

- A graph of a cumulative distribution, called an **ogive,** shows data values on the horizontal axis and either the **cumulative frequencies**, the **cumulative relative frequencies**, or the **cumulative percentage frequencies** on the vertical axis.

- Figure 4.4 illustrates an ogive for the cumulative frequencies of the peoples' age data. The ogive is constructed by plotting a point corresponding to the cumulative frequency of each class. Because the classes for the peoples' age data are 9-18, 19-28, 29-38 and so on, one-unit gaps appear from 18 to 19, 28 to 29, and so on. These gaps are eliminated by plotting points halfway between the class limits. Thus, 18.5 is used for the 9–18 class, 28.5 is used for the 19–28 class, and so on.

- The "less than or equal to 18" class with a cumulative frequency of 7 is shown on the ogive in Figure 4.4 by the point located at 18.5 on the horizontal axis and 7 on the vertical axis.

Figure 4. 4 Ogive for the peoples' age data
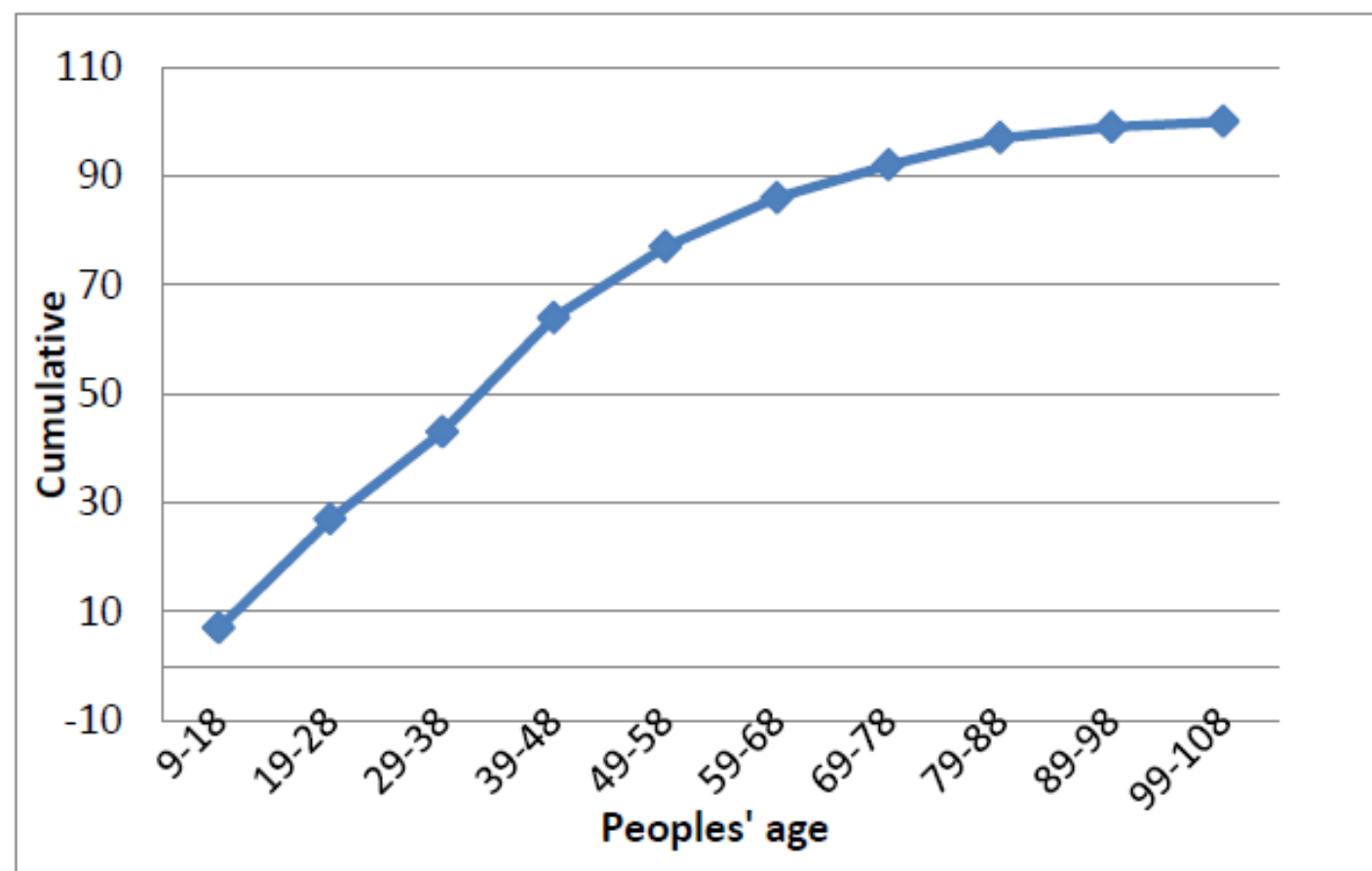
Exercises

1. Consider the following frequency distribution.

Class	Frequency
10–19	10
20–29	14
30–39	17
40–49	7
50–59	2

☐ Use Excel to answer the questions below.

1. Construct a cumulative frequency distribution and a cumulative relative frequency distribution.

2. Construct a histogram and an ogive for the data in exercise 1.

3. A doctor's office staff studied the waiting times for patients who arrive at the office with a request for emergency service. The following data with waiting times in minutes were collected over a one-month period.

2   5   10   12   4   4   5   17   11   8   9   8   12   21   6   8   7   13   18   3

In Excel, use classes of 0–4, 5–9, and so on in the following:

a. Show the frequency distribution.

b. Show the relative frequency distribution.

c. Show the cumulative frequency distribution.

d. Show the cumulative relative frequency distribution.

e. What proportion of patients needing emergency service wait 9 minutes or less?

4. A shortage of candidates has required school districts to pay higher salaries and offer extras to attract and retain school district superintendents. The following data show the annual base salary ($1000s) for superintendents in 20 districts in the greater Rochester, New York, area (*The Rochester Democrat and Chronicle, February 10, 2008*).

187	184	174	185
175	172	202	197
165	208	215	164
162	172	182	156
172	175	170	183

In Excel, use classes of 150–159, 160–169, and so on in the following.

a. Show the frequency distribution.

b. Show the percent frequency distribution.

c. Show the cumulative frequency and cumulative percent frequency distribution.

d. Construct an ogive and histogram.

e. What percentage of the superintendents make more than $200,000?

5. The Dow Jones Industrial Average (DJIA) underwent one of its infrequent reshufflings of companies when General Motors and Citigroup were replaced by Cisco Systems and Travelers (*The Wall Street Journal, June 8, 2009*). At the time, the prices per share for the 30 companies in the DJIA were as follows:

$/Share	$/Share
61	107
11	16
25	35
24	56
12	27
52	59
38	26
69	22
20	14
49	53
27	43
72	56
14	29
37	51
24	25

a. Using a class width of 10, develop a frequency distribution for the data in Excel.

b. Prepare a histogram. Discuss the general shape of the histogram and the most frequent price range.

c. Do the data appear to be skewed? Explain.

d. Show the cumulative frequency and cumulative relative frequency distribution and ogive in Excel.

Additional materials

https://www.youtube.com/watch?v=IVhQTAF1guc

https://www.youtube.com/watch?v=cEGVeJPO5Js

Thank you for attention!