ÇAĞ UNIVERSITY
FACULTY OF ECONOMICS & ADMINISTRATIVE SCIENCES

STATISTICS FOR BUSINESS
LECTURE NOTES 3

Dr. Elma Satrovic|
elmasa1991@gmail.com

Example (qualitative/categorical data)

- Table 1 summarizes the data from a sample of 50 soft drink purchases. Use Excel to develop frequency distribution (COUNTIF function), relative frequency, percentage frequency, pie chart (for relative frequency) and bar chart (for frequency).

Coke Classic	Sprite	Pepsi
Diet Coke	Coke Classic	Coke Classic
Pepsi	Diet Coke	Coke Classic
Diet Coke	Coke Classic	Coke Classic
Coke Classic	Diet Coke	Pepsi
Coke Classic	Coke Classic	Dr. Pepper
Dr. Pepper	Sprite	Coke Classic
Diet Coke	Pepsi	Diet Coke
Pepsi	Coke Classic	Pepsi
Pepsi	Coke Classic	Pepsi
Coke Classic	Coke Classic	Pepsi
Dr. Pepper	Pepsi	Pepsi
Sprite	Coke Classic	Coke Classic
Coke Classic	Sprite	Dr. Pepper
Diet Coke	Dr. Pepper	Pepsi
Coke Classic	Pepsi	Sprite
Coke Classic	Diet Coke	

Summarizing Quantitative Data

Tabular Summary

1. Frequency Distribution

- As defined in previous lecture, a frequency distribution is a tabular summary of data showing the number (frequency) of items in each of several nonoverlapping classes. This definition holds for quantitative as well as qualitative data.
- However, with quantitative data we must be more careful in defining the nonoverlapping classes to be used in the frequency distribution.
- For quantitative data, we need to define the classes first. The three steps necessary to define the classes for a frequency distribution with quantitative data are:

    1. Determine the number of nonoverlapping classes.

    2. Determine the width of each class.

    3. Determine the class limits.

- We demonstrate these steps by constructing a frequency distribution for the people's age data in Table 3.1.

Table 3 1 People's age in a small area in Wolverhampton

33	64	35	80	37	33	45	70	51	35
52	42	33	19	68	50	12	9	37	14
21	53	57	45	37	30	63	62	46	34
56	41	26	44	41	22	24	28	20	18
19	26	28	62	64	58	70	72	69	71
42	48	48	57	55	56	60	25	46	100
44	30	22	52	39	42	38	43	51	17
19	44	38	27	28	45	44	39	29	80
82	16	18	29	48	53	61	70	82	90
91	87	64	47	30	25	21	22	23	19

Step 1:

- **Number of classes -** Classes are formed by specifying ranges that will be used to group the data. As a general guideline, we recommend using between 5 and 20 classes.
  - For a small number of data items, as few as five or six classes may be used to summarize the data.
  - For a larger number of data items, a larger number of classes is usually required.
  - The goal is to use enough classes to show the variation in the data, but not so many classes that some contain only a few data items.
- Because the number of data items in Table 3.1 is relatively large *(n =100),* we chose to construct a frequency distribution with ten classes.

Step 2:

- **Width of the classes** - As a general guideline, it is recommended that the width be the same for each class. Thus the choices of the number of classes and the width of classes are not independent decisions. A larger number of classes means a smaller class width, and vice versa.
- To determine an approximate class width, we begin by identifying the largest and smallest data values. Then, with the desired number of classes specified, we can use the following expression to determine the approximate class width.

$$\text{Approximate class width} = \frac{\text{Largest data value} - \text{Smallest data value}}{\text{Number of classes}} \qquad (3.1)$$

- Please note that the approximate class width can be rounded to a more convenient value. For example, an approximate class width of 5.73 might be rounded to 6. For the data presented in Table 3.1, the largest data value is 100 and the smallest data value is 9. So the approximate class width is:

$$\text{Approximate class width} = \frac{100 - 9}{10} = \frac{91}{10} = 9.1$$

- We decided to round it up and use a class width of 10 in the frequency distribution. In practice, the number of classes and the appropriate class width are determined by trial and error.
- Once a possible number of classes are chosen, equation (3.1) is used to find the approximate class width. It, however, is up to the analyst judgment after determining multiple combinations of numbers.
- For the people's age data presented in Table 3.1, we decided using ten classes, each with a width of ten.

Step 3:

- **Class limits -** Class limits must be chosen so that each data item belongs to one and only one class.
  - The *lower class limit* identifies the smallest possible data value assigned to the class.
  - The *upper class limit* identifies the largest possible data value assigned to the class.
- In developing frequency distributions for qualitative data, we did not need to specify class limits because each data item naturally fell into a separate class. But with quantitative data, such as the people's age data presented in Table 3.1, class limits are necessary to determine where each data value belongs.

Table 3 1 People's age in a small area in Wolverhampton

33	64	35	80	37	33	45	70	51	35
52	42	33	19	68	50	12	9	37	14
21	53	57	45	37	30	63	62	46	34
56	41	26	44	41	22	24	28	20	18
19	26	28	62	64	58	70	72	69	71
42	48	48	57	55	56	60	25	46	100
44	30	22	52	39	42	38	43	51	17
19	44	38	27	28	45	44	39	29	80
82	16	18	29	48	53	61	70	82	90
91	87	64	47	30	25	21	22	23	19

- Using the age data, we selected 9 years as the lower class limit and 18 years as the upper class limit for the first class (since class width of 10 is used).
- The smallest data value, 9, is included in the 9-18 class. We then selected 19 year as the lower class limit and 28 year as the upper class limit of the next class. We continued defining the lower and upper class limits to obtain a total of ten classes: 9-18, 19-28, 29-38, 39-48, 49-58, 59-68, 69-78, 79-88, 89-98, 99-108. The largest value, 100, is included in the 99-108 class.

- The difference between the lower class limits of adjacent classes is the **class width.** Using the first two lower class limits of 9 and 19, we see that the class width is 19-9=10.
- With the number of classes, class width, and class limits determined, a frequency distribution can be obtained by counting the number of data values belonging to each class.
- For example, the data in Table 3.2 show that seven values belong to the 9–18 class. Thus, the frequency for the 9–18 class is 7. Continuing this counting process for the other classes provides the frequency distribution in Table 3.2.

Table 3 2 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

- Using the frequency distribution in Table 3.2, we can observe that:
  - The most frequent age interval is 39-48, with the frequency of 21. Twenty one out of 100 people's age are in this class.
  - It is followed by 19-28 with the frequency of 20 meaning that 20 out of 100 people's age are in this class.
  - The least frequent age interval is 99-108, with the frequency of 1 (only one person is 99 years old/ or older).

Table 3 2 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

- The value of a frequency distribution is that it provides insights about the data that are not easily obtained by viewing the data in their original unorganized form.

- **Class midpoint -** In some applications, we want to know the midpoints of the classes in a frequency distribution for quantitative data. The **class midpoint** is the value halfway between the lower and upper class limits.

$$\text{class midpoint} = \frac{\text{lower class} + \text{upper class}}{2}$$

- For the age data, the ten class midpoints are: 13.5 ((9+18)/2), 23.5, 33.5, 43.5, 53.5, 63.5, 73.5, 83.5, 93.5, and 103.5.

2. Relative Frequency and Percent Frequency Distributions

- We define the relative frequency and percent frequency distributions for quantitative data in the same manner as for qualitative data. First, recall that the **relative frequency** is the proportion of the observations belonging to a class. With *n* observations:

$$\text{Relative frequency of class} = \frac{\text{Frequency of the class}}{n}$$

- The **percent frequency** of a class is the relative frequency multiplied by 100.
- Based on the class frequencies in Table 3.2 and with *n=100,* Table 3.3 shows the *relative frequency distribution* and *percent frequency distribution* for the people's age data.

Table 3. 3 Relative and Percentage frequency distributions for the people's ages

Age (years)	Relative Frequency	Percentage Frequency	Meaning
9-18	$\frac{7}{100} = 0.07$	0.07x100=7	7 percent of the data
19-28	$\frac{20}{100} = 0.2$	0.2x100=20	20 percent of the data
29-38	$\frac{16}{100} = 0.16$	0.16x100=16	16 percent of the data
39-48	$\frac{21}{100} = 0.21$	0.21x100=21	21 percent of the data
49-58	$\frac{13}{100} = 0.13$	0.13x100=13	13 percent of the data
59-68	$\frac{9}{100} = 0.09$	0.09x100=9	9 percent of the data
69-78	$\frac{6}{100} = 0.06$	0.06x100=6	6 percent of the data
79-88	$\frac{5}{100} = 0.05$	0.05x100=5	5 percent of the data
89-98	$\frac{2}{100} = 0.02$	0.02x100=2	2 percent of the data
99-108	$\frac{1}{100} = 0.01$	0.01x100=1	1 percent of the data

- Note that 0.21 of the people, or 21%, is from 39 to 48 years old. Only 0.01 of the people, or 1%, is 99 years old/ or older. Again, additional interpretations and insights can be obtained by using Table 3.3.

Excel application 1

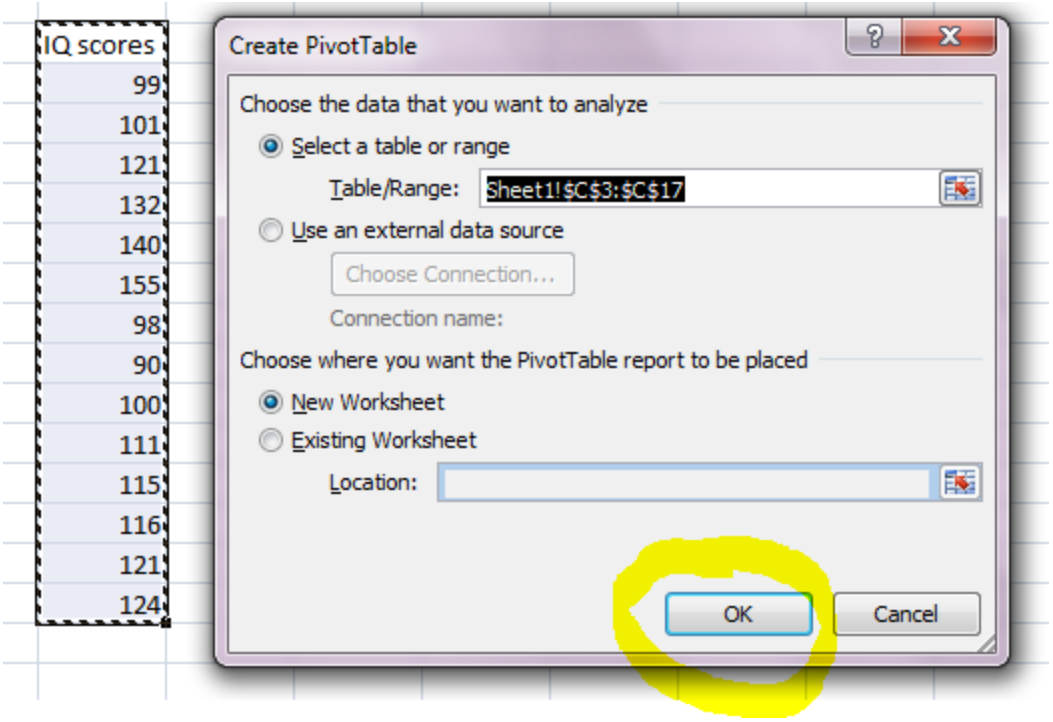
- **Sample Problem:** Make a frequency distribution table in Excel by using pivot table. Use the following IQ scores: 99, 101, 121, 132, 140, 155, 98, 90, 100, 111, 115, 116, 121, 124 (class width is 10).

Solution

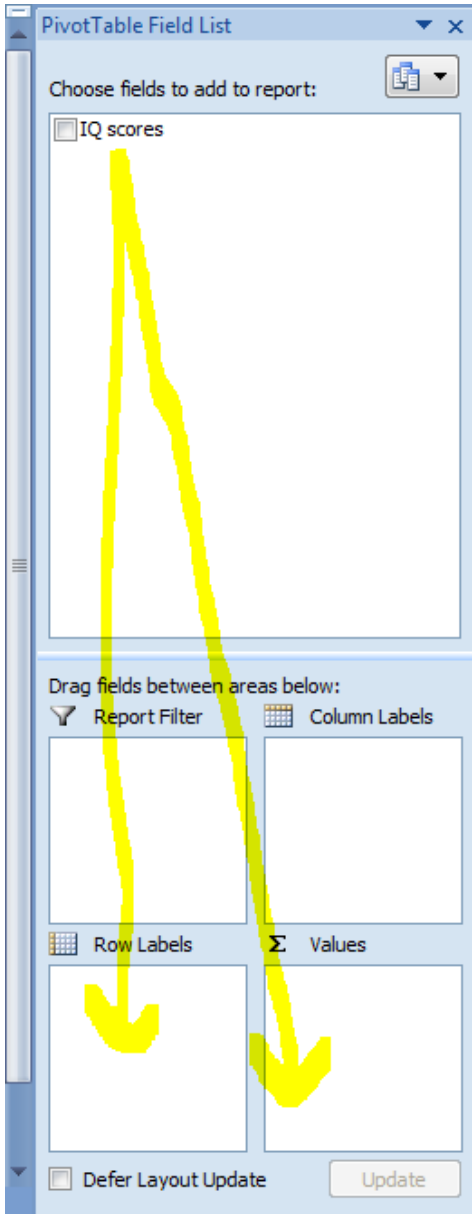
Step 1

IQ scores
99
101
121
132
140
155
98
90
100
111
115
116
121
124

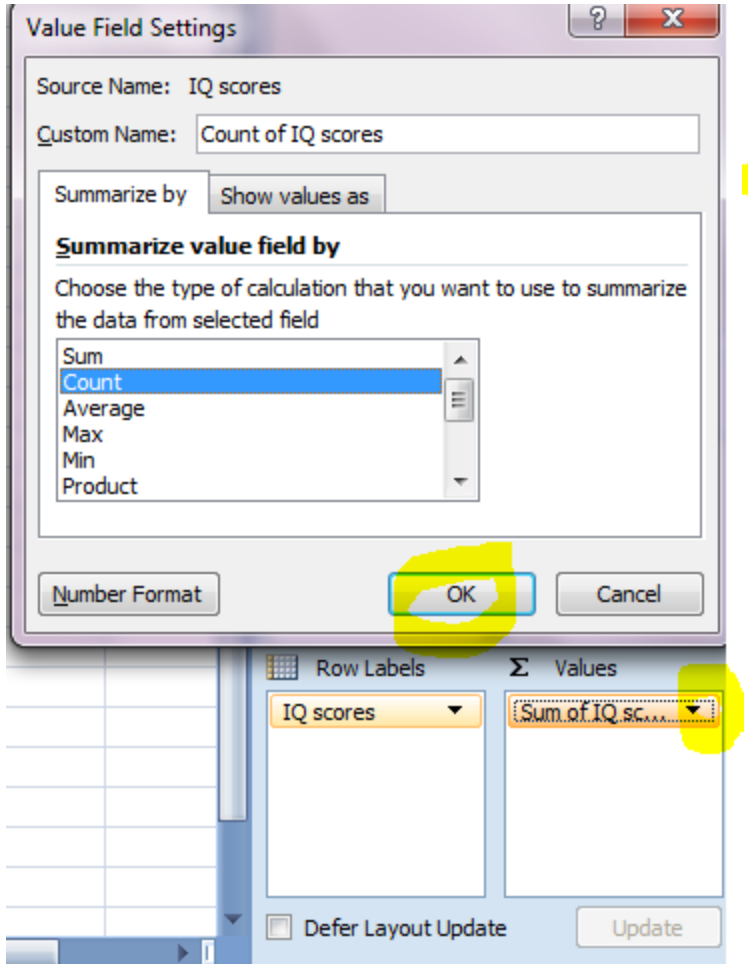
Step 2



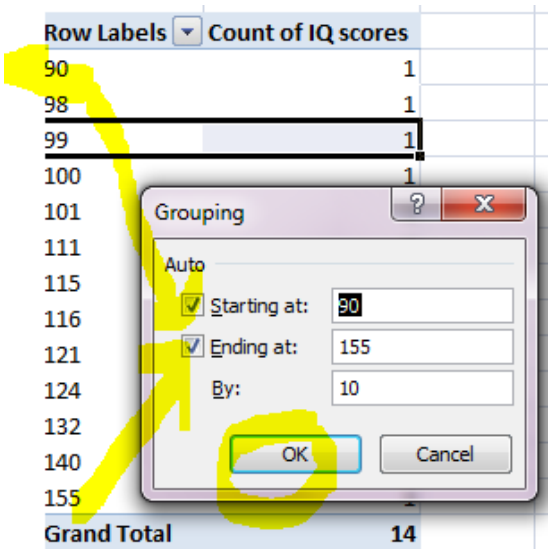
Step 3



Step 4



Step 5



Step 6

Row Labels	Count of IQ scores
90-99	3
100-109	2
110-119	3
120-129	3
130-139	1
140-149	1
150-159	1
Grand Total	14

Excel application 2

- Based on the data from the Table 1, use pivot table to develop a frequency distribution in Excel considering that class width is 3.
- Table 1

14	21	23	21	16
19	22	25	16	16
24	24	25	19	16
19	18	19	21	12
16	17	18	23	25
20	23	16	20	19
24	26	15	22	24
20	22	24	22	20

Solution

Row Labels	Count of Data
12-14	2
15-17	8
18-20	11
21-23	10
24-26	9
Grand Total	**40**

Exercises

1. Consider the following data. Use Excel to answer the questions below (class width is 2).

8.9	10.2	11.5	7.8	10.0
6.8	9.5	11.5	11.2	14.9
12.2	13.5	14.1	10.0	12.2
7.5	10.0	6.0	15.8	11.5

a. Construct a frequency distribution.

b. Construct a percent frequency distribution.

2. Consider the following data. Use Excel to answer the questions below.

14	21	23	21	16
19	22	25	16	16
24	24	25	19	16
19	18	19	21	12
16	17	18	23	25
20	23	16	20	19
24	26	15	22	24
20	22	24	22	20

a. Develop a frequency distribution using classes of 12–15, 16–19, 20–23, 24–27.

b. Develop a relative frequency distribution and a percent frequency distribution using the classes in part (a).

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. Use Excel to answer the questions below.

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

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. What is the number of patients needing emergency service wait between 15 and 19 minutes?

d. What percentage of patients needing emergency service wait between 5 and 9 minutes? (percentage frequency)

e. What proportion of patients needing emergency service wait between 5 and 9 minutes? (relative frequency)

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

Use classes of 150–159, 160–169, and so on in the following.

a. Show the frequency distribution in Excel.

b. Show the percent frequency distribution in Excel.

c. What percentage of the superintendents make between $200,000 and $209,000?

d. How many of the of the superintendents make the lowest salary?

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. What is the highest price per share? What is the lowest price per share?

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

c. Show the relative frequency distribution in Excel.

d. Show the percent frequency distribution in Excel.

Thank you for attention!