

Statistical Analysis

Some of the slides in this presentation are from Bowerman, B. L., O'Connell, R. T., & Murphree, E.S., (2010). *Business Statistics in Practice* (6th Ed.), Copyright © McGraw-Hill Education.

McGraw-Hill makes no representations or warranties as to the accuracy of any information contained in the McGraw-Hill Material, including any warranties of merchantability or fitness for a particular purpose. In no event shall McGraw-Hill have any liability to any party for special, incidental, tort, or consequential damages arising out of or in connection with the McGraw-Hill Material, even if McGraw-Hill has been advised of the possibility of such damages.

Descriptive Statistics: Tabular and Graphical Methods

1. Graphically Summarizing Qualitative Data
2. Graphically Summarizing Quantitative Data

Graphically Summarizing Qualitative Data

With qualitative data, names identify the different categories

The data can be summarized using a frequency distribution

Frequency distribution: A table that summarizes the number of items in each of several non-overlapping classes

Bowerman, B. L., O'Connell, R. T., & Murphree, E. S., (2010). Business Statistics in Practice (6th Ed.), Copyright © McGraw-Hill Education

Bar Charts and Pie Charts

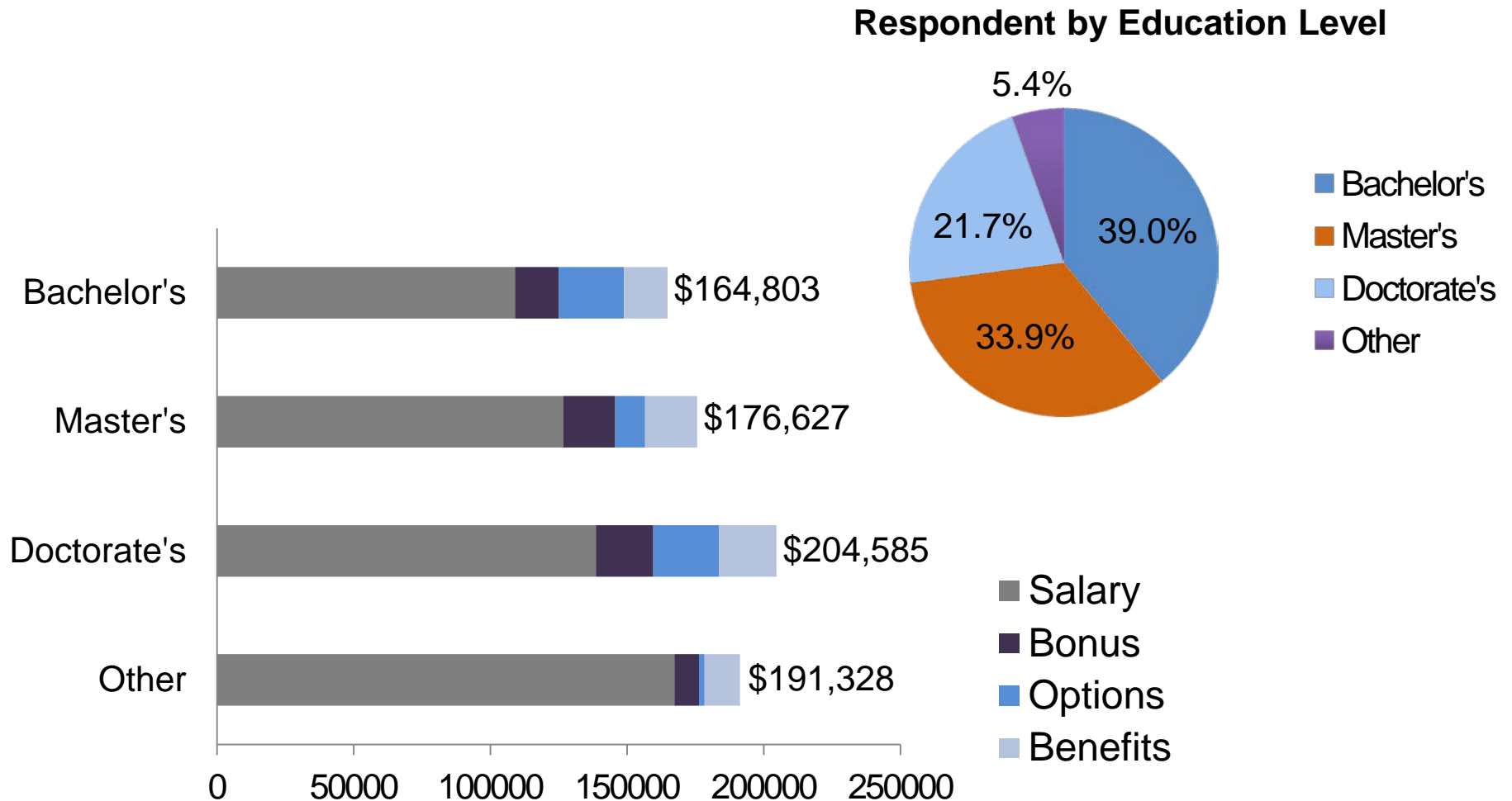
Bar chart: A vertical or horizontal rectangle represents the frequency for each category

Height can be frequency, relative frequency, or percent frequency

Pie chart: A circle divided into slices where the size of each slice represents its relative frequency or percent frequency

Bowerman, B. L., O'Connell, R. T., & Murphree, E. S., (2010). Business Statistics in Practice (6th Ed.), Copyright © McGraw-Hill Education

Bar and Pie Chart of Education Based Compensation Level



Adapted from Contract Pharma. "13th Annual Salary Survey" June 2012: 48. Print.

Graphically Summarizing Qualitative Data

Often need to summarize and describe the shape of the distribution

One way is to group the measurements into classes of a frequency distribution and then displaying the data in the form of a histogram

A *frequency distribution* is a list of data classes with the count of values that belong to each class

- “Classify and count”
- The frequency distribution is a table

Bowerman, B. L., O'Connell, R. T., & Murphree, E. S., (2010). Business Statistics in Practice (6th Ed.), Copyright © McGraw-Hill Education

Constructing a Frequency Distribution

Steps in making a frequency distribution:

1. Find the number of classes
2. Find the class length
3. Form non-overlapping classes of equal width
4. Tally and count
5. Graph the histogram

Bowerman, B. L., O'Connell, R. T., & Murphree, E. S., (2010). Business Statistics in Practice (6th Ed.), Copyright © McGraw-Hill Education

Frequency Distribution *example*

Raw data for the number of patrons at the new art museum per week day ($n=50$)

68	71	77	83	79	72	74	57	67	69
60	70	66	76	70	84	59	75	94	65
85	79	71	83	84	74	82	97	77	50
93	95	78	81	79	90	83	84	101	72
86	93	92	102	80	69	80	91	73	78

Number of Classes

Group all of the n data into c number of classes
 c is the smallest whole number for which $2^c \geq n$

In Examples $n=50$

- For $c=5$, $2^5 = 32, < n$
- For $c=6$, $2^6 = 64, > n$
- So use $c=6$ classes

Adapted from Bowerman, B. L., O'Connell, R. T., & Murphree, E. S., (2010). Business Statistics in Practice (6th Ed.), Copyright © McGraw-Hill Education

Class Length

Find the length of each class as the largest measurement minus the smallest divided by the number of classes found earlier (c)

For Example, $(102-50)/6 = 8.677$

- Round to ten

68	71	77	83	79	72	74	57	67	69
60	70	66	76	70	84	59	75	94	65
85	79	71	83	84	74	82	97	77	50
93	95	78	81	79	90	83	84	101	72
86	93	92	102	80	69	80	91	73	78

Adapted from Bowerman, B. L., O'Connell, R. T., & Murphree, E. S., (2010). Business Statistics in Practice (6th Ed.), Copyright © McGraw-Hill Education

Form Non-Overlapping Classes of Equal Width

The classes start on the smallest value

- This is the lower limit of the first class

The upper limit of the first class is smallest value + class length

- In the example, the first class starts at 50 people and goes up to 59

The next class starts at this upper limit, 60 and goes up by class length

And so on

Adapted from Bowerman, B. L., O'Connell, R. T., & Murphree, E. S., (2010). Business Statistics in Practice (6th Ed.), Copyright © McGraw-Hill Education

Tally and Count

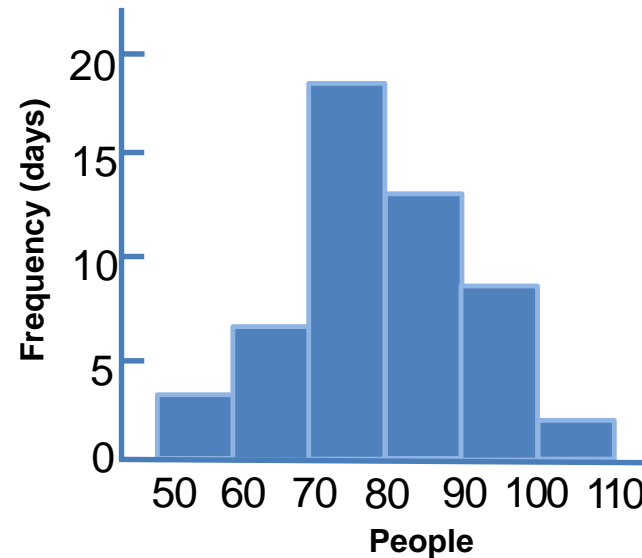
Class (people)	Frequency (days)
50 to 59	3
60 to 69	7
70 to 79	18
80 to 89	12
90 to 99	8
100 to 109	2
	50

68	71	77	83	79	72	74	57	67	69
60	70	66	76	70	84	59	75	94	65
85	79	71	83	84	74	82	97	77	50
93	95	78	81	79	90	83	84	101	72
86	93	92	102	80	69	80	91	73	78

Histogram

- Rectangles represent the classes
- The base represents the class length
- The height represents
 - a) the frequency in a frequency histogram, or
 - b) the relative frequency in a relative frequency histogram

Class (people)	Frequency (days)
50 to 59	3
60 to 69	7
70 to 79	18
80 to 89	12
90 to 99	8
100 to 109	2
	50



Adapted from Bowerman, B. L., O'Connell, R. T., & Murphree, E. S., (2010). Business Statistics in Practice (6th Ed.), Copyright © McGraw-Hill Education

Some Common Distribution Shapes

Skewed to the right: The right tail of the histogram is longer than the left tail

Skewed to the left: The left tail of the histogram is longer than the right tail

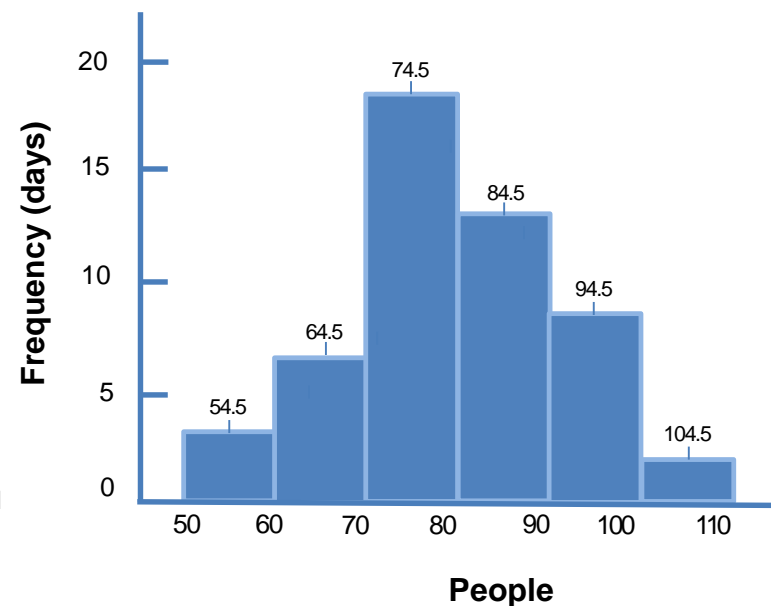
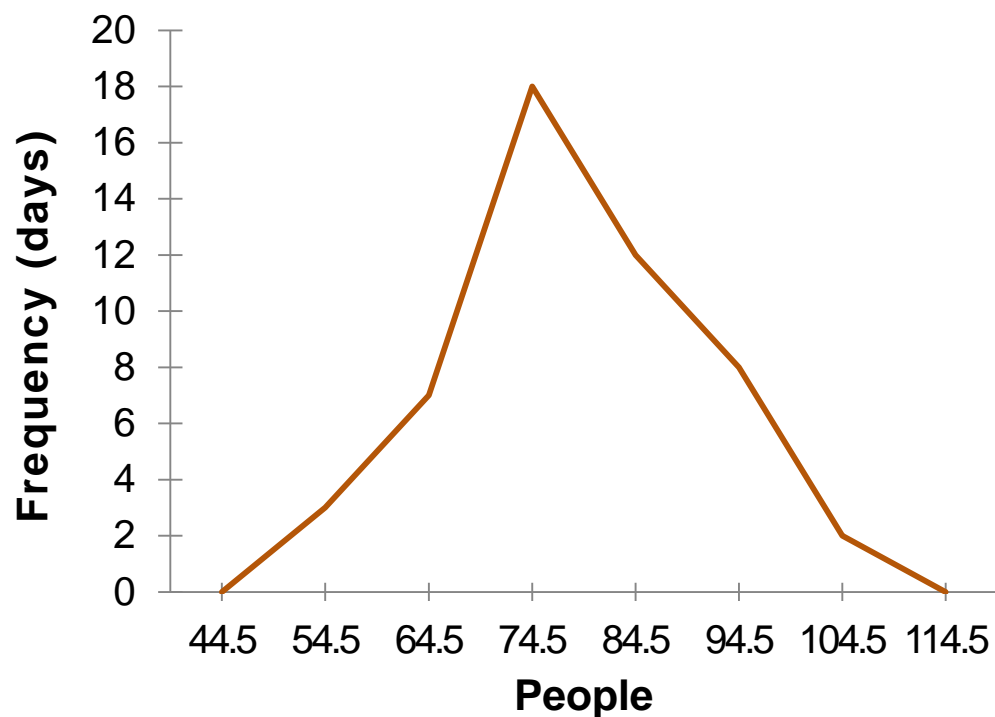
Symmetrical: The right and left tails of the histogram appear to be mirror images of each other

Bowerman, B. L., O'Connell, R. T., & Murphree, E. S., (2010). Business Statistics in Practice (6th Ed.), Copyright © McGraw-Hill Education

Frequency Polygons

Plot a point above each class midpoint at a height equal to the frequency of the class

Useful when comparing two or more distributions



Adapted from Bowerman, B. L., O'Connell, R. T., & Murphree, E. S., (2010). Business Statistics in Practice (6th Ed.), Copyright © McGraw-Hill Education

Cumulative Distributions

Another way to summarize a distribution is to construct a cumulative distribution

To do this, use the same number of classes, class lengths, and class boundaries used for the frequency distribution

Rather than a count, we record the number of measurements that are less than the upper boundary of that class

In other words, a running total

Bowerman, B. L., O'Connell, R. T., & Murphree, E. S., (2010). Business Statistics in Practice (6th Ed.), Copyright © McGraw-Hill Education

Less Than Cumulative Frequency Distribution

Class (people)	Frequency (days)
50 to 59	3
60 to 69	7
70 to 79	18
80 to 89	12
90 to 99	8
100 to 109	2
	50

Class (people)	Cumulative Frequency (days)
Less than 50	0
Less than 60	3
Less than 70	10
Less than 80	28
Less than 90	40
Less than 100	48
Less than 110	50

Relative Greater Than Cumulative Frequency Distribution

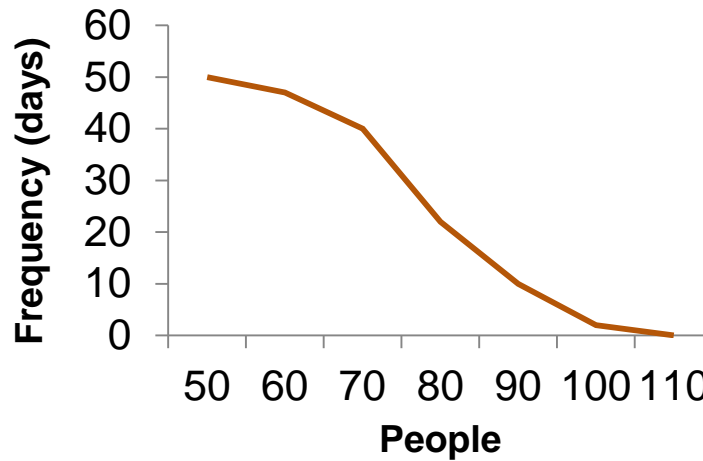
Class (people)	Frequency (days)
50 to 59	3
60 to 69	7
70 to 79	18
80 to 89	12
90 to 99	8
100 to 109	2
	50

Class (people)	Cumulative Frequency (days)	Relative Cumulative Frequency (days)
50 or more	50	100% (=50/50)
60 or more	47	94% (=47/50)
70 or more	40	80% (=40/50)
80 or more	22	44% (=22/50)
90 or more	10	20% (=10/50)
100 or more	2	4% (=2/50)
110 or more	0	0% (=0/50)

Ogive

A graph of a cumulative distribution

- Plot a point above each upper class boundary at height of cumulative frequency
- Connect points with line segments
- Can also be drawn using
 - a) Cumulative relative frequencies
 - b) Cumulative percent frequencies



Adapted from Bowerman, B. L., O'Connell, R. T., & Murphree, E. S., (2010). Business Statistics in Practice (6th Ed.), Copyright © McGraw-Hill Education