Organizing Data; Histograms, Frequency Polygons, and Ogives

2-1 and 2-2

Jacob Ayers

MAT 110

Lesson #3

Objectives

- Organize data using categorical frequency distributions, ungrouped frequency distributions, and grouped frequency distributions
- Find class width, boundaries, and midpoints for grouped data
- Use histograms, frequency polygons, and ogives to visualize frequency distributions

Organizing Data

Last week, you learned how researchers gather data.

Organizing Data

Last week, you learned how researchers gather data.

Now, we'll turn our attention to organizing that data.

Organizing Data

Last week, you learned how researchers gather data.

Now, we'll turn our attention to organizing that data.

Definition

A *frequency distribution* is the organization of raw data in table form, using classes and frequencies.

Organizing Data

Last week, you learned how researchers gather data.

Now, we'll turn our attention to organizing that data.

Definition

A *frequency distribution* is the organization of raw data in table form, using classes and frequencies.

When creating a frequency distribution, we divide the data into categories (classes) and then determine how many data values fall into each category (frequencies).

Categorical Frequency Distributions

When data can be placed into qualitative categories, we use categorical frequency distibutions.

Categorical Frequency Distributions

When data can be placed into qualitative categories, we use categorical frequency distibutions.

Examples:

- blood type
- political party
- college major

Categorical Frequency Distributions

When data can be placed into qualitative categories, we use categorical frequency distibutions.

Examples:

- blood type
- political party
- college major

To make the frequency distribution, tally how many data values are in each category and list it in table form. We can also determine relative frequencies, using percentages.

Categorical Frequency Distributions

Twenty-five army inductees were given a blood test to determine their blood type. The data set is below. Construct a categorical frequency distribution for the data.

Categorical Frequency Distributions

Twenty-five army inductees were given a blood test to determine their blood type. The data set is below. Construct a categorical frequency distribution for the data.

A	B	B	AB	O
O	O	B	AB	B
B	B	O	A	O
A	O	O	O	AB
AB	A	O	B	A

Categorical Frequency Distributions

Twenty-five army inductees were given a blood test to determine their blood type. The data set is below. Construct a categorical frequency distribution for the data.

A	B	B	AB	O
O	O	B	AB	B
B	B	O	A	O
A	O	O	O	AB
AB	A	O	B	A

Blood Type	Frequency	Percent
A	5	$5 \div 25 = 20\%$
B	7	$7 \div 25 = 28\%$
O	9	$9 \div 25 = 36\%$
AB	4	$4 \div 25 = 16\%$
TOTALS	25	100%

Ungrouped Frequency Distributions

When dealing with numerical data that is all in a small range, an ungrouped frequency distribution can be used.

Ungrouped Frequency Distributions

When dealing with numerical data that is all in a small range, an ungrouped frequency distribution can be used.

An ungrouped frequency distribution is just a categorical distribution where the categories are quantitative rather than qualitative.

Ungrouped Frequency Distributions

When dealing with numerical data that is all in a small range, an ungrouped frequency distribution can be used.

An ungrouped frequency distribution is just a categorical distribution where the categories are quantitative rather than qualitative.

Example: The ages of 20 dogs in a pet shelter are shown. Construct an ungrouped frequency distribution for the data.

Ungrouped Frequency Distributions

When dealing with numerical data that is all in a small range, an ungrouped frequency distribution can be used.

An ungrouped frequency distribution is just a categorical distribution where the categories are quantitative rather than qualitative.

Example: The ages of 20 dogs in a pet shelter are shown. Construct an ungrouped frequency distribution for the data.

5	8	7	6	3
9	4	4	5	8
7	4	7	5	7
3	5	8	4	9

Ungrouped Frequency Distributions

When dealing with numerical data that is all in a small range, an ungrouped frequency distribution can be used.

An ungrouped frequency distribution is just a categorical distribution where the categories are quantitative rather than qualitative.

Example: The ages of 20 dogs in a pet shelter are shown. Construct an ungrouped frequency distribution for the data.

5	8	7	6	3
9	4	4	5	8
7	4	7	5	7
3	5	8	4	9

Since there are only 7 unique ages, it makes sense to use an ungrouped distribution here.

Ungrouped Frequency Distributions

Here is the ungrouped frequency distribution:

Age	Frequency
3	2
4	4
5	4
6	1
7	4
8	3
9	2

Grouped Frequency Distributions

When data values are spread over a large range, it doesn't make sense to have individual classes for each value. We must group data into wider classes.

Grouped Frequency Distributions

When data values are spread over a large range, it doesn't make sense to have individual classes for each value. We must group data into wider classes.

Below is an example of a grouped frequency distribution:

Class limits	Class boundaries	Tally	Frequency
58–64	57.5–64.5	/	1
65–71	64.5–71.5	𝍸 /	6
72–78	71.5–78.5	𝍸 𝍸	10
79–85	78.5–85.5	𝍸 𝍸 ////	14
86–92	85.5–92.5	𝍸 𝍸 //	12
93–99	92.5–99.5	𝍸	5
100–106	99.5–106.5	//	2
			Total 50

Grouped Frequency Distributions

Here is the first row again:

Grouped Frequency Distributions

Here is the first row again:

Class limits	Class boundaries	Frequency
58-64	57.5-64.5	1

Grouped Frequency Distributions

Here is the first row again:

Class limits	Class boundaries	Frequency
58-64	57.5-64.5	1

Some things to note:

- 58 = lower class limit - smallest data value in the class

Grouped Frequency Distributions

Here is the first row again:

Class limits	Class boundaries	Frequency
58-64	57.5-64.5	1

Some things to note:

- 58 = lower class limit - smallest data value in the class
- 64 = upper class limit - largest data value in the class

Grouped Frequency Distributions

Here is the first row again:

Class limits	Class boundaries	Frequency
58-64	57.5-64.5	1

Some things to note:

- 58 = lower class limit - smallest data value in the class
- 64 = upper class limit - largest data value in the class
- Class boundaries - eliminate gaps in the distribution

Grouped Frequency Distributions

Here is the first row again:

Class limits	Class boundaries	Frequency
58-64	57.5-64.5	1

Some things to note:

- 58 = lower class limit - smallest data value in the class
- 64 = upper class limit - largest data value in the class
- Class boundaries - eliminate gaps in the distribution
- Class width = (Lower Limit) - (Lower Limit) (in this case, 65 - 58 = 7)

Grouped Frequency Distributions

Some rules to follow when constructing a grouped frequency distribution:

Grouped Frequency Distributions

Some rules to follow when constructing a grouped frequency distribution:

- Use 5-20 classes

Grouped Frequency Distributions

Some rules to follow when constructing a grouped frequency distribution:

- Use 5-20 classes
- Preferably, use odd class width (ensures midpoints are same place value as data)

Grouped Frequency Distributions

Some rules to follow when constructing a grouped frequency distribution:

- Use 5-20 classes
- Preferably, use odd class width (ensures midpoints are same place value as data)
- Classes must be *mutually exclusive* (i.e. no overlap)

Grouped Frequency Distributions

Some rules to follow when constructing a grouped frequency distribution:

- Use 5-20 classes
- Preferably, use odd class width (ensures midpoints are same place value as data)
- Classes must be *mutually exclusive* (i.e. no overlap)
- Classes must be *continuous* (i.e. no gaps)

Grouped Frequency Distributions

Some rules to follow when constructing a grouped frequency distribution:

- Use 5-20 classes
- Preferably, use odd class width (ensures midpoints are same place value as data)
- Classes must be *mutually exclusive* (i.e. no overlap)
- Classes must be *continuous* (i.e. no gaps)
- Classes must be *exhaustive* (i.e. all data values must be in a class)

Grouped Frequency Distributions

Some rules to follow when constructing a grouped frequency distribution:

- Use 5-20 classes
- Preferably, use odd class width (ensures midpoints are same place value as data)
- Classes must be *mutually exclusive* (i.e. no overlap)
- Classes must be *continuous* (i.e. no gaps)
- Classes must be *exhaustive* (i.e. all data values must be in a class)
- Classes must be equal in width

Grouped Frequency Distributions

Procedure Table

Constructing a Grouped Frequency Distribution

Step 1 Determine the classes.

- Find the highest and lowest values.
- Find the range.
- Select the number of classes desired.
- Find the width by dividing the range by the number of classes and rounding up.
- Select a starting point (usually the lowest value or any convenient number less than the lowest value); add the width to get the lower limits.
- Find the upper class limits.
- Find the boundaries.

Step 2 Tally the data.

Step 3 Find the numerical frequencies from the tallies, and find the cumulative frequencies.

Grouped Frequency Distributions

The number of stories in each of a sample of the world's 30 tallest buildings follows. Construct a grouped frequency distribution with 7 classes.

88	88	110	88	80	69	102	78	70	55
79	85	80	100	60	90	77	55	75	55
54	60	75	64	105	56	71	70	65	72

Grouped Frequency Distributions

The number of stories in each of a sample of the world's 30 tallest buildings follows. Construct a grouped frequency distribution with 7 classes.

88	88	110	88	80	69	102	78	70	55
79	85	80	100	60	90	77	55	75	55
54	60	75	64	105	56	71	70	65	72

Step 1: Determine the Classes

Highest Value: 110; Lowest Values: 54

Grouped Frequency Distributions

The number of stories in each of a sample of the world's 30 tallest buildings follows. Construct a grouped frequency distribution with 7 classes.

88	88	110	88	80	69	102	78	70	55
79	85	80	100	60	90	77	55	75	55
54	60	75	64	105	56	71	70	65	72

Step 1: Determine the Classes

Highest Value: 110; Lowest Values: 54

Range = 110 - 54 = 56

Grouped Frequency Distributions

The number of stories in each of a sample of the world's 30 tallest buildings follows. Construct a grouped frequency distribution with 7 classes.

88	88	110	88	80	69	102	78	70	55
79	85	80	100	60	90	77	55	75	55
54	60	75	64	105	56	71	70	65	72

Step 1: Determine the Classes

Highest Value: 110; Lowest Values: 54

Range = 110 - 54 = 56

Class width = $56 \div 7 = 8$; round up to next odd number: 9

Grouped Frequency Distributions

The number of stories in each of a sample of the world's 30 tallest buildings follows. Construct a grouped frequency distribution with 7 classes.

88	88	110	88	80	69	102	78	70	55
79	85	80	100	60	90	77	55	75	55
54	60	75	64	105	56	71	70	65	72

Step 1: Determine the Classes

Highest Value: 110; Lowest Values: 54

Range = 110 - 54 = 56

Class width = $56 \div 7 = 8$; round up to next odd number: 9

Starting point: 54

Grouped Frequency Distributions

88	88	110	88	80	69	102	78	70	55
79	85	80	100	60	90	77	55	75	55
54	60	75	64	105	56	71	70	65	72

Steps 2 and 3: Tally Data; Find Numerical Frequencies

Class Limits	Class Boundaries	Tally	Frequency
54-62			
63-71			
72-80			
81-89			
90-98			
99-107			
108-116			

Grouped Frequency Distributions

88	88	110	88	80	69	102	78	70	55
79	85	80	100	60	90	77	55	75	55
54	60	75	64	105	56	71	70	65	72

Steps 2 and 3: Tally Data; Find Numerical Frequencies

Class Limits	Class Boundaries	Tally	Frequency
54-62	53.5-62.5		
63-71	62.5-71.5		
72-80	71.5-80.5		
81-89	80.5-89.5		
90-98	89.5-98.5		
99-107	98.5-107.5		
108-116	107.5-116.5		

Grouped Frequency Distributions

88	88	110	88	80	69	102	78	70	55
79	85	80	100	60	90	77	55	75	55
54	60	75	64	105	56	71	70	65	72

Steps 2 and 3: Tally Data; Find Numerical Frequencies

Class Limits	Class Boundaries	Tally	Frequency
54-62	53.5-62.5		7
63-71	62.5-71.5		6
72-80	71.5-80.5		8
81-89	80.5-89.5		4
90-98	89.5-98.5		1
99-107	98.5-107.5		3
108-116	107.5-116.5		1

Histograms, Frequency Polygons, and Ogives

Once data has been organized, it is often helpful to display it visually.

Histograms, Frequency Polygons, and Ogives

Once data has been organized, it is often helpful to display it visually.

The three most common graphs in research are:

- Histograms
- Frequency Polygons
- Ogives

Histograms, Frequency Polygons, and Ogives

Once data has been organized, it is often helpful to display it visually.

The three most common graphs in research are:

- Histograms
- Frequency Polygons
- Ogives

We will go over each graph in turn.

Histograms, Frequency Polygons, and Ogives

Below is the procedure for constructing these three graphs.

Below is the procedure for constructing these three graphs.

Procedure Table

Constructing a Histogram, Frequency Polygon, and Ogive

Step 1	Draw and label the x and y axes.
Step 2	On the x axis, label the class boundaries of the frequency distribution for the histogram and ogive. Label the midpoints for the frequency polygon.
Step 3	Plot the frequencies for each class, and draw the vertical bars for the histogram and the lines for the frequency polygon and ogive.

(*Note:* Remember that the lines for the frequency polygon begin and end on the x axis while the lines for the ogive begin on the x axis.)

Below is the procedure for constructing these three graphs.

Procedure Table

Constructing a Histogram, Frequency Polygon, and Ogive

Step 1	Draw and label the x and y axes.
Step 2	On the x axis, label the class boundaries of the frequency distribution for the histogram and ogive. Label the midpoints for the frequency polygon.
Step 3	Plot the frequencies for each class, and draw the vertical bars for the histogram and the lines for the frequency polygon and ogive.

(*Note:* Remember that the lines for the frequency polygon begin and end on the x axis while the lines for the ogive begin on the x axis.)

We will use Google Sheets to construct these graphs.

Histograms

Histogram: displays data using vertical bars

Histograms

Histogram: displays data using vertical bars

Construct a histogram for the following data:

Class boundaries	Frequency
99.5-104.5	2
104.5-109.5	8
109.5-114.5	18
114.5-119.5	13
119.5-124.5	7
124.5-129.5	1
129.5-134.5	1

Frequency Polygons

Frequency Polygon: displays data using lines that connect points plotted for frequencies at midpoints of classes

Frequency Polygons

Frequency Polygon: displays data using lines that connect points plotted for frequencies at midpoints of classes

Construct a frequency polygon for the following data:

Class boundaries	Midpoints	Frequency
99.5-104.5		2
104.5-109.5		8
109.5-114.5		18
114.5-119.5		13
119.5-124.5		7
124.5-129.5		1
129.5-134.5		1

Ogives

Ogive: line graph that represents cumulative frequencies for each class

Ogives

Ogive: line graph that represents cumulative frequencies for each class

Construct an ogive for the following data:

Class boundaries	Frequency
99.5-104.5	2
104.5-109.5	8
109.5-114.5	18
114.5-119.5	13
119.5-124.5	7
124.5-129.5	1
129.5-134.5	1

	Cumulative Frequency
Less than 99.5	0
Less than 104.5	2
Less than 109.5	10
Less than 114.5	28
Less than 119.5	41
Less than 124.5	48
Less than 129.5	49
Less than 134.5	50

Putting It Together

The number of faculty listed for a sample of private colleges that offer only bachelor's degrees is listed below. Use these data to construct a frequency distribution with 7 classes, a histogram, a frequency polygon, and an ogive.

Putting It Together

The number of faculty listed for a sample of private colleges that offer only bachelor's degrees is listed below. Use these data to construct a frequency distribution with 7 classes, a histogram, a frequency polygon, and an ogive.

165	221	218	206	138	135	224	204
70	210	207	154	155	82	120	116
176	162	225	214	93	389	77	135
221	161	128	310				

Step 1: Determine the Classes

Highest: 389; Lowest: 70

Putting It Together

The number of faculty listed for a sample of private colleges that offer only bachelor's degrees is listed below. Use these data to construct a frequency distribution with 7 classes, a histogram, a frequency polygon, and an ogive.

165	221	218	206	138	135	224	204
70	210	207	154	155	82	120	116
176	162	225	214	93	389	77	135
221	161	128	310				

Step 1: Determine the Classes

Highest: 389; Lowest: 70

Range: 389 - 70 = 329

Putting It Together

The number of faculty listed for a sample of private colleges that offer only bachelor's degrees is listed below. Use these data to construct a frequency distribution with 7 classes, a histogram, a frequency polygon, and an ogive.

165	221	218	206	138	135	224	204
70	210	207	154	155	82	120	116
176	162	225	214	93	389	77	135
221	161	128	310				

Step 1: Determine the Classes

Highest: 389; Lowest: 70

Range: 389 - 70 = 329

Class width = $329 \div 7 = 47$

Putting It Together

The number of faculty listed for a sample of private colleges that offer only bachelor's degrees is listed below. Use these data to construct a frequency distribution with 7 classes, a histogram, a frequency polygon, and an ogive.

165	221	218	206	138	135	224	204
70	210	207	154	155	82	120	116
176	162	225	214	93	389	77	135
221	161	128	310				

Step 1: Determine the Classes

Highest: 389; Lowest: 70

Range: 389 - 70 = 329

Class width = $329 \div 7 = 47$

Starting point: 70

Putting It Together

165	221	218	206	138	135	224	204
70	210	207	154	155	82	120	116
176	162	225	214	93	389	77	135
221	161	128	310				

Steps 2 and 3: Tally Data; Find Numerical Frequencies

Class Limits	Class Boundaries	Class Midpoints	Tally	Frequency
70-116				
117-163				
164-210				
211-257				
258-304				
305-351				
352-398				

Putting It Together

165	221	218	206	138	135	224	204
70	210	207	154	155	82	120	116
176	162	225	214	93	389	77	135
221	161	128	310				

Steps 2 and 3: Tally Data; Find Numerical Frequencies

Class Limits	Class Boundaries	Class Midpoints	Tally	Frequency
70-116	69.5-116.5			
117-163	116.5-163.5			
164-210	163.5-210.5			
211-257	210.5-257.5			
258-304	257.5-304.5			
305-351	304.5-351.5			
352-398	351.5-398.5			

Putting It Together

165	221	218	206	138	135	224	204
70	210	207	154	155	82	120	116
176	162	225	214	93	389	77	135
221	161	128	310				

Steps 2 and 3: Tally Data; Find Numerical Frequencies

Class Limits	Class Boundaries	Class Midpoints	Tally	Frequency
70-116	69.5-116.5	93		
117-163	116.5-163.5	140		
164-210	163.5-210.5	187		
211-257	210.5-257.5	234		
258-304	257.5-304.5	281		
305-351	304.5-351.5	328		
352-398	351.5-398.5	375		

Putting It Together

165	221	218	206	138	135	224	204
70	210	207	154	155	82	120	116
176	162	225	214	93	389	77	135
221	161	128	310				

Steps 2 and 3: Tally Data; Find Numerical Frequencies

Class Limits	Class Boundaries	Class Midpoints	Tally	Frequency
70-116	69.5-116.5	93		5
117-163	116.5-163.5	140		9
164-210	163.5-210.5	187		6
211-257	210.5-257.5	234		6
258-304	257.5-304.5	281		0
305-351	304.5-351.5	328		1
352-398	351.5-398.5	375		1

Next Steps

- Read 2-3
- Watch Video Lesson #4
- Complete Assignment 2

Thanks for watching!