

Frequency Tables and Histograms

Purpose:

Participants will identify intervals for a data set, create a frequency table that describes the data in terms of these intervals, and generate a corresponding histogram for displaying the data.

Overview:

In small groups, participants will study a data set, decide on reasonable intervals for subdividing the data, create a table which includes frequency and relative frequency of each interval, and generate the corresponding histogram for displaying the data.

TEXES Mathematics 4-8 Competencies. The beginning teacher:

- IV.012.A Organizes and displays data in a variety of formats (e.g., tables, frequency distributions, stem-and-leaf plots, box-and-whisker plots, histograms, pie charts).
- IV.012.B Applies concepts of center, spread, shape, and skewness to describe a data distribution.
- IV.012.C Supports arguments, makes predictions, and draws conclusions using summary statistics and graphs to analyze and interpret one-variable data.
- IV.012.E Analyzes connections among concepts of center and spread, data clusters and gaps, data outliers, and measures of central tendency and dispersion.
- IV.012.F Calculates and interprets percentiles and quartiles.

TEKS Mathematics Objectives. The student is expected to:

- 5.13.C Graph a given set of data using an appropriate graphical representation such as a picture or line.
- 6.10.A Draw and compare different graphical representations of the same data.
- 7.11.A Select and use an appropriate representation for presenting collected data and justify the selection.
- 8.12.C Construct circle graphs, bar graphs, and histograms, with and without technology.

Terms.

Spread of data, mean, class intervals, bins, frequency, relative frequency, cumulative relative frequency, histogram

Materials.

For instructor:

- Transparencies
- Overhead graphing calculator and LCD panel

For each participant:

- Data Sheet
- Activity Sheet
- Graphing calculator

Transparencies.

- *Frequency Tables and Histograms*

Procedure:

Steps	Questions/Math Notes
<p>1. Have participants enter the data either by hand or by linking two calculators and sending the lists from the calculator that has the data to the other calculator.</p> <p>Ask participants to work in groups of 4 and to check one another's input of data.</p>	<p>Let L1 = Region Number, L2 = Total Number of Students Enrolled in that Region during 1987-88, and L3 = Total Number of Students Enrolled in that Region during 1997-98.</p> <p>Do not enter the 4th column of data.</p>
<p>2. Once the data for L1, L2, and L3 are entered into all calculators, have participants compute the total student enrollment change from 1987-88 to 1997-98 in L4.</p> <p>Once L4 has been calculated, have participants exchange calculators with their partner and check the L4 column for correctness. If L4 is correct, then most likely all the data is correct.</p>	<p>Calculating the total student enrollment change from 1987-88 to 1997-98 in L4 can be done by placing the cursor on L4 and typing L3 – L2. Then press ENTER.</p>
<p>3. Ask participants to decide on the number of intervals they would like to use for separating the data listed in L3. Then they are to determine the interval size (bin width) needed.</p>	<p><i>How many class intervals are reasonable for this data? What happens to the data as the number of class intervals increases? decreases?</i></p> <p><i>How do we determine the bin width once we decide the number of intervals?</i></p> <p>(We decide which values we are going to use for Xmin and Xmax. Then we subtract Xmin from Xmax and divide the difference by the number of class intervals.)</p>
<p>4. Ask participants to make a frequency table that shows the class intervals (or bins), the frequency of data points for each bin, the relative frequency, and the cumulative relative frequency.</p>	<p><i>What are the class intervals that you have chosen to use?</i></p> <p>(Answers will vary.)</p> <p><i>How do we determine the frequency of data points for each class interval?</i></p> <p>(We count the number of data points which lie in each class interval.)</p> <p><i>How do we determine the relative frequency?</i></p> <p>(The relative frequency for a class interval is equal to the frequency divided by the total number of data points.)</p> <p><i>How do we determine the cumulative relative</i></p>

	<p><i>frequency?</i></p> <p>(The cumulative relative frequency for a class interval is equal to the sum of the frequencies that lie in the interval plus all the prior intervals divided by the total number of data points.)</p>
<p>5. Ask participants to sketch a histogram of the data using the information from the frequency table they created in step #4.</p> <p>Ask participants to create a histogram of the data on their graphing calculator. (Refer to the Calculator Instructions below.)</p>	<p><i>What does the histogram tell us about the data in L3 that wasn't as obvious from the measures of central tendency and variability?</i></p> <p>(The histogram shows us that most of the education service regions enrolled less than 200,000 students and that the two regions with more than 500,000 students (549,212 students in Region X and 828,262 in Region IV) are definite outliers.)</p>

Calculator Instructions.

Developing Class Intervals for a Data Set. The data in L3 spans from 42,388 to 828,262. We could choose 9 intervals (bins) which span from 0 to 900,000 with each bin width equal to 100,000. Alternatively, we could choose 8 intervals (bins) which span from 40,000 to 840,000 again with each bin width equal to 100,000. Also, we could choose 6 intervals (bins) which span from 20,000 to 860,000 with each bin width equal to 140,000, etc.

Suppose we choose 10 intervals (bins) which span from 40,000 to 840,000 with each bin width equal to 80,000. Then the 1st interval would be 40,000 – 120,000, the 2nd interval would be 120,000 – 200,000, etc.

Creating the Frequency Table.

Class Interval	Bin Width	Frequency	Relative Frequency	Cumulative Relative Frequency
40,000 to < 120,000	80,000	10	0.50	0.50
120,000 to < 200,000	80,000	04	0.20	0.70
200,000 to < 280,000	80,000	01	0.05	0.75
280,000 to < 360,000	80,000	02	0.10	0.85
360,000 to < 440,000	80,000	01	0.05	0.90
440,000 to < 520,000	80,000	00	0.00	0.90
520,000 to < 600,000	80,000	01	0.05	0.95
600,000 to < 680,000	80,000	00	0.00	0.95
680,000 to < 760,000	80,000	00	0.00	0.95
760,000 to < 840,000	80,000	01	0.05	1.00

Setting Up the WINDOW for Class (Bin Width).

Set the WINDOW in the following manner:

- X min = 40000 (smallest value of 1st bin)
- X max = 840000 (largest value of 10th bin)
- X scl = 80000 (bin width)
- Y min = -3 (minimum frequency of data points in interval)
- Y max = 12 (maximum frequency of data points in interval)
- Y scl = 0 (this places no scale markings on vertical axis)
- X res = 1

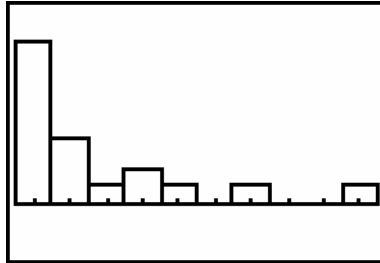
```

WINDOW
Xmin=40000
Xmax=840000
Xscl=80000
Ymin=-3
Ymax=12
Yscl=0
Xres=1

```

Y min has very little meaning in this exercise so letting $Y \text{ min} = -(Y \text{ max} \div 4)$ leaves enough room at the bottom of the plot screen for the class (bin) information.

Setting Up the Plot. To graph, turn on STAT PLOT 1, choose the 3rd plot (Histogram), input x list: L3, Freq: 1, and press GRAPH. Press TRACE and use the ► and ◀ keys to move from interval to interval. Note that the height of each interval bar represents the frequency.



What does the histogram tell us about the data in L3 that wasn't as obvious from the measures of central tendency and variability?

Determining the Effect of Changes to the Interval Width or Xscl.

In the WINDOW, change the Xscl from 80000 to 40000. What effect does this change have on the graph?

Change the Xmin to 0, the Xmax to 900000, the Xscl to 300000, and the Ymax to 20.

Have each person in the group choose a different Xscl and compare graphs. What effect do these changes have on the graph?

Solutions to Activity Sheet:

1. Answers will vary. An example answer is provided in the Calculator Instructions under Developing Class Intervals.
2. An example frequency table is shown in the Calculator Instructions under Creating the Frequency Table.
3. An example histogram is shown in the Calculator Instructions under Setting Up the Plot.
4. The histogram shows us that most of the regions enrolled less than 200,000 students and that the two regions with more than 500,000 students (549,212 students in Region X and 828,262 in Region IV) are definite outliers.