

# Box-and-Whisker Plots

**Purpose:**

Participants will create regular and modify box-and-whisker plots using two data sets.

**Overview:**

Pairs of participants will generate regular box-and-whisker plots of L2 and L3. Using the TRACE feature of the graphing calculator, they will investigate each value that occurs as they trace each boxplot. Then, they will generate modified box-and-whisker plots of L2 and L3 and compare these graphs to the regular box-and-whisker plots.

**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.D Demonstrates an understanding of measures of central tendency (e.g., mean, median, mode) and dispersion (e.g., range, interquartile range, variance, standard deviation).
- 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.B Describe characteristics of data presented in tables and graphs including the shape and spread of the data and the middle number.
- 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.
- 6.10.B Use median, mode, and range to describe data.
- 7.11.B Make inferences and convincing arguments based on an analysis of given or collected data.
- 7.12.A Describe a set of data using mean, median, mode, and range.
- 7.12.B Choose among mean, median, mode, or range to describe a set of data and justify the choice for a particular situation.
- 8.13.B Recognize misuses of graphical or numerical information and evaluate predictions and conclusions based on data analysis.

**Terms.**

Data, minimum and maximum values, median, 1<sup>st</sup> and 3<sup>rd</sup> quartiles, range, interquartile range, box-and-whisker plot

**Materials.**

For instructor:

- Transparencies
- Overhead graphing calculator and LCD panel

For each participant:

- Data Sheet
- Activity Sheet
- Graphing calculator

**Transparencies.**

- *Box-and-Whisker Plots*

**Activity Sheet(s).**

- *Box-and-Whisker Plots*

**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 which has the data to the other calculator.</p> <p>Ask participants to work in pairs 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 4<sup>th</sup> 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 generate a regular Boxplot (the fifth type) of the L2 data in STAT PLOT 1 and graph. Then trace the Boxplot.</p> <p>Ask participants what WINDOW should be used to be able to see the graph.</p>	<p><i>What values occur as you trace the regular Boxplot? What do these values represent?</i> (The values that occur for the L2 data are: minX = 40,517 students – 1<sup>st</sup> data point Q1 = 66,997 students – 25<sup>th</sup> percentile Med = 104,697 students – 50<sup>th</sup> percentile Q3 = 201,580.5 students – 75<sup>th</sup> percentile maxX = 659,516 students – 20<sup>th</sup> data point)</p>
<p>4. Ask participants to generate a modified Boxplot (the fourth type) of the L2 data in STAT PLOT 2 and graph. Then trace the modified Boxplot.</p> <p>IQR stands for interquartile range. IQR = Q3 – Q1.</p>	<p><i>What values occur as you trace the modified Boxplot? What do these values represent?</i> (The values that occur for the L2 data are: minX = 40,517 students – 1<sup>st</sup> data point Q1 = 66,997 students – 25<sup>th</sup> percentile Med = 104,697 students – 50<sup>th</sup> percentile Q3 = 201,580.5 students – 75<sup>th</sup> percentile X = 286,784 students – 18<sup>th</sup> data point X = 426,631 students – 19<sup>th</sup> data point maxX = 659,516 students – 20<sup>th</sup> data point)</p> <p>(The modified Boxplot draws the whiskers from the 1<sup>st</sup> and 3<sup>rd</sup> quartiles either to the most extreme data value in that direction or to a distance of 1.5 times the IQR, whichever is closer. In this case, the left whisker is drawn to data point 1 but the right whisker is drawn to data point 18. This indicates that data points 19 and 20 are outliers in the data.)</p>

<p>5. Ask participants to compare the regular Boxplot to the modified Boxplot.</p>	<p><i>How does the modified box-and-whisker plot differ from the regular box-and-whisker plot?</i></p> <p>(The right whisker is shorter on the modified Boxplot. Instead of plotting the whiskers to the minimum and maximum data points as is done with the regular Boxplot, the calculator graphs the whiskers to the most extreme data points that lie no more than 1.5 x IQR distance from the quartiles. On the left side of the Boxplot, the minimum data point lies within this range; on the right side the maximum data point does not lie within the range. The outliers in the data--data points 19 and 20--become apparent with the modified Boxplot.)</p>
<p>6. Ask participants to set up STAT PLOT 3 for a histogram of L2.</p> <p>Use the window (0, 900000, 100000, -3, 15, 0) for graphing. STAT PLOT 1, STAT PLOT 2, and STAT 3 should all appear in the same window of the calculator.</p>	<p><i>How does the information provided by the histogram differ from the boxplots?</i></p> <p>(The histogram provides information about the frequency of data within intervals. In this case the intervals are [0-100,000), [100,000-200,000), [200,000-300,000), [300,000-400,000), etc. It gives us a feel for how the data is distributed. The boxplots, on the otherhand, provide different information about how the data is distributed. They provide the values of the minimum and maximum values and the 25<sup>th</sup>, 50<sup>th</sup>, and 75<sup>th</sup> percentiles of the data. The modified boxplot also indicates data points that are outliers.)</p>
<p>7. Ask participants to turn off STAT PLOT 1 and create a modified Boxplot of L3 in STAT PLOT 3. They are to view STAT PLOT 2 and STAT PLOT 3 on the same screen.</p>	<p><i>How do the modified Boxplots of L2 and L3 differ?</i></p> <p>(The values that occur for the L2 data are:  minX = 40,517 students – 1<sup>st</sup> data point  Q1 = 66,997 students – 25<sup>th</sup> percentile  Med = 104,697 students – 50<sup>th</sup> percentile  Q3 = 201,580.5 students – 75<sup>th</sup> percentile  X = 286,784 students – 18<sup>th</sup> data point  X = 426,631 students – 19<sup>th</sup> data point  maxX = 659,516 students – 20<sup>th</sup> data point)</p> <p>(The values that occur for the L3 data are:  minX = 42,388 students – 1<sup>st</sup> data point  Q1 = 69,220.5 students – 25<sup>th</sup> percentile  Med = 120,286 students – 50<sup>th</sup> percentile  Q3 = 266,301.5 students – 75<sup>th</sup> percentile  X = 549,212 students – 19<sup>th</sup> data point  maxX = 828,262 students – 20<sup>th</sup> data point)</p>

	(The box part of the L3 data is wider, indicating that the middle 50% of the data is more dispersed than that of L2. The whisker on the right side is much longer on the L3 data than the L2 data, indicating that the maximum value of the L3 data is larger than the maximum value of the L2 data.)
--	---

**Calculator Instructions.**

**Turning on the Statplot.** To get to the Stat Plot screen, press: 2<sup>nd</sup>, Y=. This will take you to STAT PLOTS. If any of the plots are turned on, turn them off by pressing 4: PlotsOff. To turn on STAT PLOT 1, press 2<sup>nd</sup>, Y=, 1:Plot1, highlight On, ENTER.

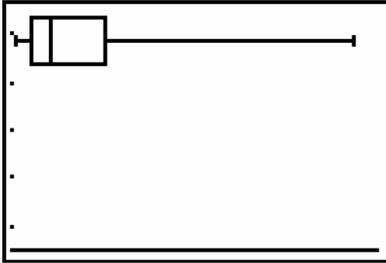
**Setting-up Statplot 1 of L2.** Turn on Stat Plot 1 by pressing 2<sup>nd</sup>, Y=, 1:Plot1, highlight On, ENTER. This will take you into Plot 1. Using the arrow key, press the down arrow key once, to move to Type. The cursor will flash on the 1<sup>st</sup> type of graph. We will be using the 5<sup>th</sup> type, so press the right arrow key until the middle graph on the 2<sup>nd</sup> row is highlighted. Then press, ENTER. (Pressing the down arrow key will not move the cursor to the second row of graph types.) Input Xlist: L2 and Freq: 1.

Before we press GRAPH, we need to modify the window so that the boxplot appears on the graph screen. To change the window, we go to WINDOW. Decide on a minimum X value that is smaller than the lowest data point, and a maximum X value that is greater than your largest data point. The X scale is up to you and how many points you want on the X-axis. For the Ymin = -3 so that the x-axis will appear, and for Ymax = 12, with a Yscl= 0, and xres=1. Now, press GRAPH to view the Boxplot.

**Solutions to Activity Sheet:**

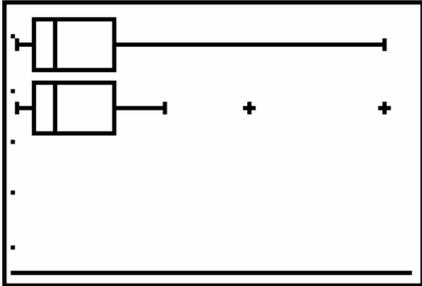
1. *What values occur as you trace the regular Boxplot? What do these values represent?*

- L2 data in Regular Boxplot
- minX = 40,517 students – 1<sup>st</sup> data point
  - Q1 = 66,997 students – 25<sup>th</sup> percentile
  - Med = 104,697 students – 50<sup>th</sup> percentile
  - Q3 = 201,580.5 students – 75<sup>th</sup> percentile
  - maxX = 659,516 students – 20<sup>th</sup> data point



2. *What values occur as you trace the modified Boxplot? What do these values represent?*

- L2 data in Modified Boxplot
- minX = 40,517 students – 1<sup>st</sup> data point
  - Q1 = 66,997 students – 25<sup>th</sup> percentile
  - Med = 104,697 students – 50<sup>th</sup> percentile
  - Q3 = 201,580.5 students – 75<sup>th</sup> percentile
  - X = 286,784 students – 18<sup>th</sup> data point
  - X = 426,631 students – 19<sup>th</sup> data point
  - maxX = 659,516 students – 20<sup>th</sup> data point



The modified Boxplot draws the whiskers from the 1<sup>st</sup> and 3<sup>rd</sup> quartiles either to the most extreme data value in that direction or to a distance of 1.5 times the IQR, whichever is closer.

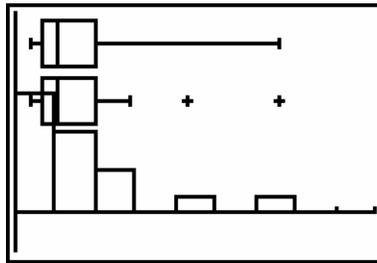
In this case, the left whisker is drawn to data point 1 but the right whisker is drawn to data point 18. This indicates that data points 19 and 20 are outliers in the data.

3. *How does the modified Boxplot differ from the regular Boxplot?*

The right whisker is shorter on the modified Boxplot. Instead of plotting the whiskers to the minimum and maximum data points as is done with the regular Boxplot, the calculator graphs the whiskers to the most extreme data points that lie no more than  $1.5 \times \text{IQR}$  distance from the quartiles. On the left side of the Boxplot, the minimum data point lies within this range; on the right side, the maximum data point does not lie within the range. The outliers in the data--data points 19 and 20--become apparent with the modified Boxplot.

4. *How does the information provided by the histogram differ from the boxplots?*

The histogram provides information about the frequency of data within intervals. In this case the intervals are [0-100,000), [100,000-200,000), [200,000-300,000), [300,000-400,000), etc. It gives us a feel for how the data is distributed. The boxplots, on the otherhand, provide different information about how the data is distributed. They provide the values of the minimum and maximum values and the 25<sup>th</sup>, 50<sup>th</sup>, and 75<sup>th</sup> percentiles of the data. The modified boxplot also indicates data points that are outliers.



5. *How do the modified Boxplots of L2 and L3 differ?*

The values that occur for the L2 data are:

minX	= 40,517 students	– 1 <sup>st</sup> data point
Q1	= 66,997 students	– 25 <sup>th</sup> percentile
Med	= 104,697 students	– 50 <sup>th</sup> percentile
Q3	= 201,580.5 students	– 75 <sup>th</sup> percentile
X	= 286,784 students	– 18 <sup>th</sup> data point
X	= 426,631 students	– 19 <sup>th</sup> data point
maxX	= 659,516 students	– 20 <sup>th</sup> data point.

The values that occur for the L3 data are:

minX	= 42,388 students	– 1 <sup>st</sup> data point
Q1	= 69,220.5 students	– 25 <sup>th</sup> percentile
Med	= 120,286 students	– 50 <sup>th</sup> percentile
Q3	= 266,301.5 students	– 75 <sup>th</sup> percentile
X	= 549,212 students	– 19 <sup>th</sup> data point
maxX	= 828,262 students	– 20 <sup>th</sup> data point.

The box part of the L3 data is wider indicating that the middle 50% of the data is more dispersed than that of L2. The whisker on the right side is much longer on the L3 data than the L2 data, indicating that the maximum value of the L3 data is larger than the maximum value of the L2 data.

