Section 1.3 – Organizing and Displaying Quantitative Data

MDM4U
Jensen

Numeric Variable: A quantitative variable that takes numerical values for which it makes sense to find an average. These variables can be either continuous or discrete

Part 1: Game of Greed

Rules:

Everyone stands. Someone throws a die twice and totals the numbers. This is everybody's current score. Those that are happy with that score sit down; they have finished this round. They record their score.

For the others, the die is rolled again. Those still standing get to add the number to their total, UNLESS it is a 2. If it is a 2, the game is over and all those standing receive 0 for that round.

Keep throwing the die until a 2 comes up or everyone has sat down and recorded their score for that round. A game consists of 5 rounds.

At the end of the game, the students add their 5 scores to get their total.

Individual Results:

Round 1	Round 2	Round 3	Round 4	Round 5

My total score is: _____

Class Results:

Part 2: Stemplots

A simple graphical display for fairly small data sets of a quantitative variable is a _____(also called a stem-and-leaf plot). We made a stemplot to display the scores for the game of greed played at the beginning of this lesson.

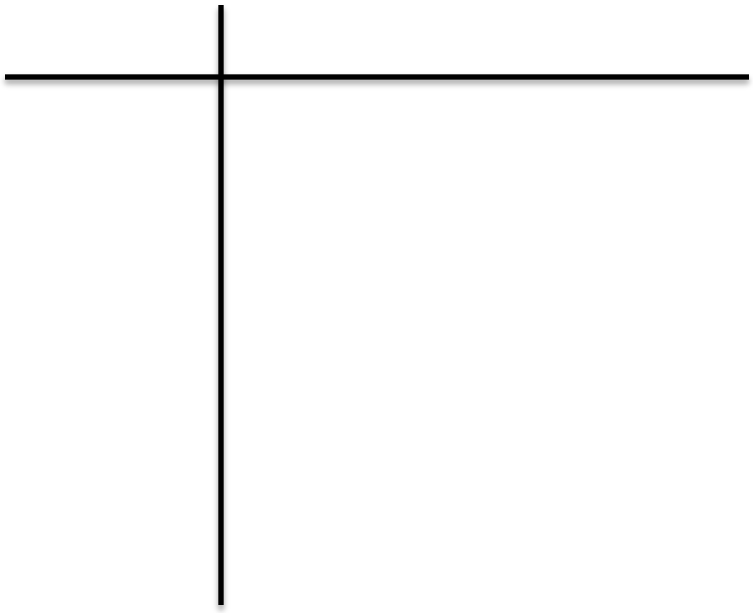
Rules for making a stemplot:

- Each number is separated into…
  - Stem: _____
  - Leaf: _____
- Write the stems in a vertical column with the smallest at the top.
  - Do NOT skip any stems, even is there is not data value for a particular stem.
- Draw a vertical line at the right of this column
- Write each leaf, in increasing order, in the row to the right of its stem

Example 1: The points for the 30 NHL teams from the 2013 regular season are recorded below:

72, 56, 55, 49, 48, 63, 62, 57
56, 48, 57, 51, 42, 40, 36, 77
60, 56, 55, 41, 59, 55, 45, 42
39, 66, 59, 57, 51, 48

Display the data using a stemplot:



Part 3: Boxplots (5 number summary)

Let's start by watching a video introducing boxplots:

http://www.learner.org/courses/againstallodds/unitpages/unit05.html

While watching the video, fill in answers to the following five questions:

1. What variable is used to compare different brands of hot dogs?

2. What name do we give to the value for which one-quarter of the data values falls at or below it?

3. What numbers make up a five-number summary?

4. How do you calculate the interquartile range?

5. Boxplots show that poultry hot dogs as a group differ from all-beef hot dogs. Compare the distribution of calories between the two types of hot dogs.

A boxplot is a _____that shows the distribution of a set of quantitative data. The five numbers a boxplot displays are:

Min: smallest data value
Q_1:
Median (Q_2):
Q_3:
Max:

Note: The median of a set of data is the _____piece of data. If there are an even number of data points in a set, the median is the _____of the two middle most pieces of data.

In a boxplot, the box contains the _____of the data and its width represents the _____of the data.

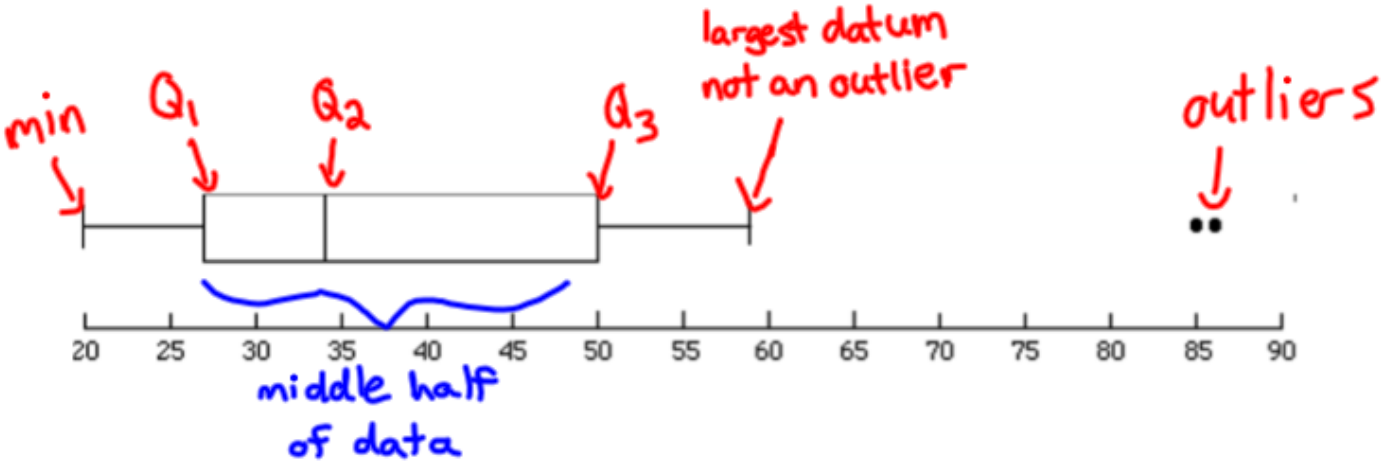
The upper and lower limits for the box are found by finding the _____for the upper and lower half of the data set. The median value itself is not included in the lower or upper half.

From the sides of the box, horizontal lines are drawn extending to the _____ and _____ data points that are NOT outliers.

_____ values are used to determine which pieces of data are _____.

Lower threshold =
Upper threshold =

Note: IQR stands for interquartile range which is =



Steps to drawing a boxplot:

- Make sure the data points are in order from smallest to largest
- Find the 5-number summary (1-Var Stats), identify outliers.
- Draw a scale (number line) above which plot will be drawn—include numbers and units: can be vertical or horizontal.
- Draw rectangular box with ends at Q_1 and Q_3.
- Draw line through box at median (Q_2).
- Draw two "whiskers" from corresponding ends of box to most extreme data value that is not an outlier—inside thresholds. Put dots or other marks for each outlier value.

Note: *Don't draw thresholds on boxplot—they are not data values. Only use them to identify outliers.*

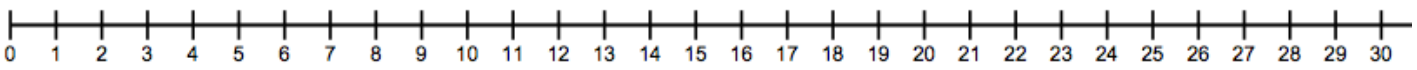
Example 2: A random survey of people at a golf course asked them how many times they had seen Happy Gilmore. The results are shown below in ascending order

1, 2, 2, 3, 3, 3, 4, 4, 5, 5, 6, 7, 8, 9, 10, 10, 12, 15, 26

a) Find the five number summary of the data (can use 1-var stats on calculator)

b) Identify outliers

c) Create a boxplot of the data



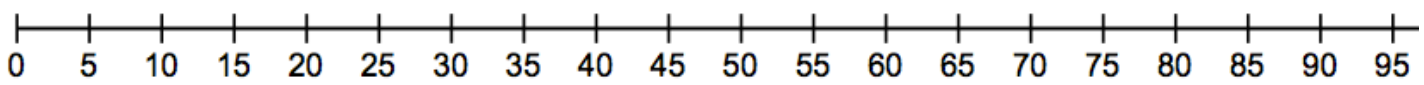
Example 3:

The times, in minutes, it took ten police officers to complete routine paperwork after their shift are given below

10, 32, 36, 38, 41, 43, 44, 48, 80, 89

Find the five-number summary all ten officers. Draw a full (modified) boxplot, showing how you identified any outliers.

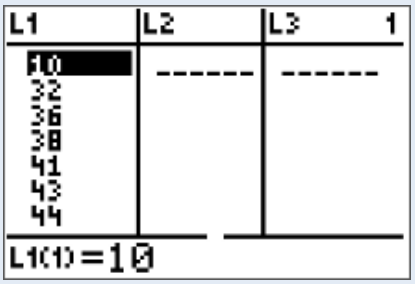
Thresholds:

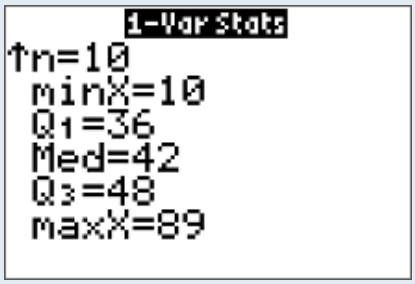


Using Ti 83/84 for Boxplots

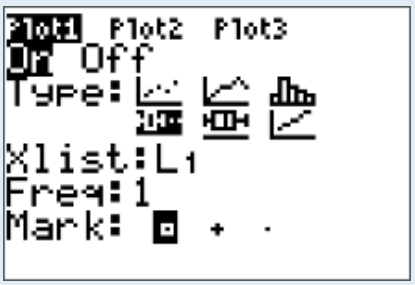
Below shows how the previous example can be completed using the ti-83/84 calculator.

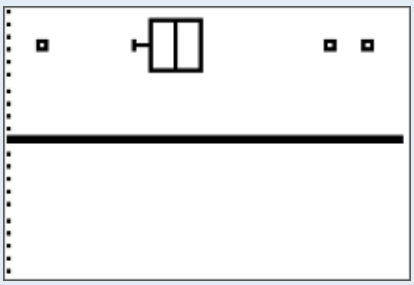
- input data in to list L1: STAT → EDIT
- Determine values for 5 number summary: STAT → CALC → 1-VARSTATS → List: L1 → CALCULATE





- turn on statplot: 2nd → y= → ENTER → ENTER
- select modified boxplot
- view graph: GRAPH → ZOOM → ZOOMSTAT



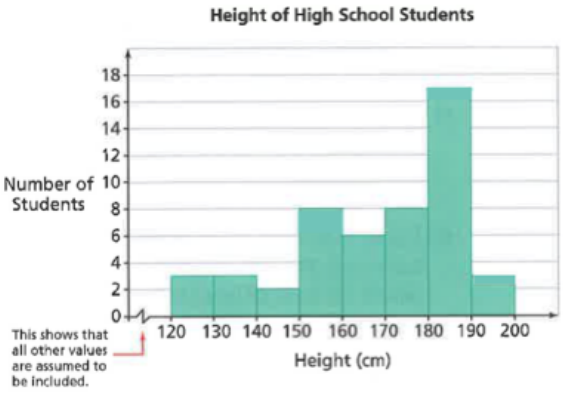


Part 4: Histograms

Quantitative variables often take many values. A graph of the distribution is clearer if nearby values are grouped together. A _____ is a frequency distribution where the horizontal access is divided into equal class _____ in to which data have been divided. The heights of the rectangles (**that have no spaces between them**) represent the frequencies associated with the corresponding intervals. It is important that each interval have the same width. Histograms are most appropriate for _____ variables but you will see them for _____ variables as well.

Example 4:

a) Is the following graph a bar graph or a histogram? How do you know?



b) Which height interval has the highest frequency? What is the frequency?

Steps for making a histogram:

1. Choose the number of intervals (if the question doesn't specify, choose between 5 and 10)

2. Calculate the range of your data (largest data point – smallest data point)

3. Round your range UP to a number that is easily divided by the number of intervals you chose.

4. Calculate your bin width; $\boldsymbol{bin\ width} = \frac{\boldsymbol{range}}{\boldsymbol{number\ of\ intervals}}$

5. Determine the first value for your first interval; $\boldsymbol{lowest\ value} - \frac{\boldsymbol{rounded\ range - actual\ range}}{\mathbf{2}}$

6. If any data points fall on the border of any of the intervals, add a decimal place to ensure that this doesn't happen.

7. Make a frequency table using the intervals you have determined.

8. Draw the histogram (no spaces between bars)

Example 5:

Here are a class' scores obtained on a data management exam:

78, 81, 55, 60, 65, 86, 44, 90

77, 71, 62, 41, 80, 72, 70, 64

88, 73, 61, 70, 75, 98, 51, 73

59, 68, 65, 81, 78, 67

a) Determine the range of the data

b) Determine an appropriate bin (interval) width that will divide the data into 6 intervals.

> Note:
>
> Round your range UP to a value that can be divided easily.

c) Determine the first value of your first interval

We added ___ to 57 when we rounded our range, therefore we should subtract _____from our smallest value _____; which makes our starting point _____.

Or use formula:

Note:

1. If you have rounded your range up you should subtract half of the amount you rounded from the smallest value to evenly distribute the 'excess of your range'.

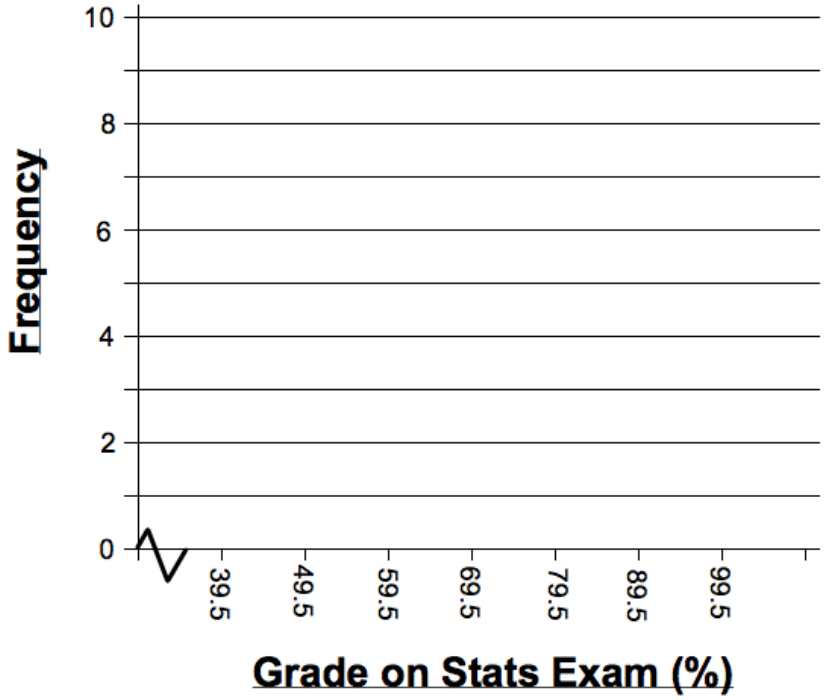
2. Make sure no data points lie on the border of two intervals. (Do this by subtracting .5 from a whole number, .05 from data with one decimal point, .005 from data with two decimal points and so on)

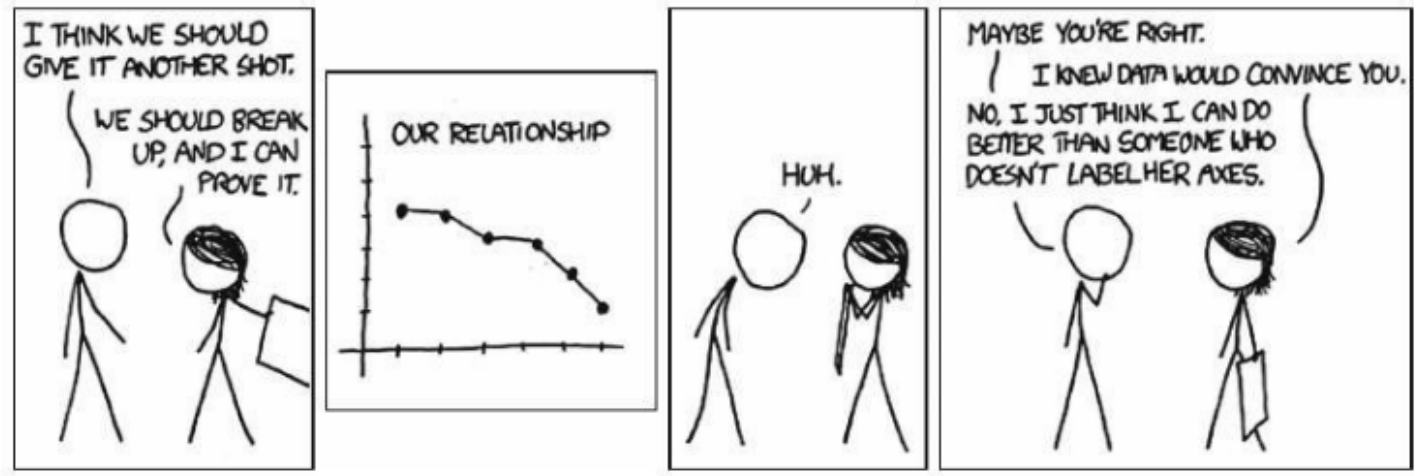
d) Create a frequency table using your intervals

Notice that the number one interval ends with, the next interval starts with the same number. This is because the data for a histogram is continuous!!!

Grade Interval	Frequency
39.5 - 49.5	
49.5 - 59.5	
59.5 - 69.5	
69.5 - 79.5	
79.5 - 89.5	
89.5 - 99.5	

e) Create a histogram of the data

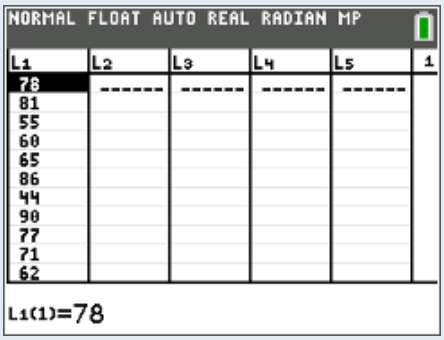


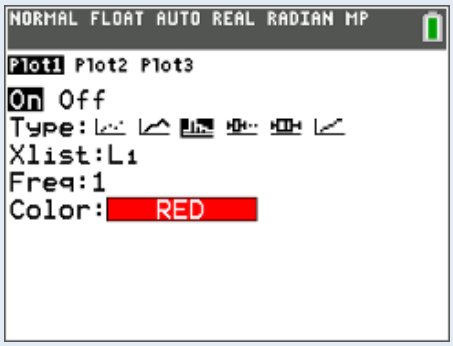


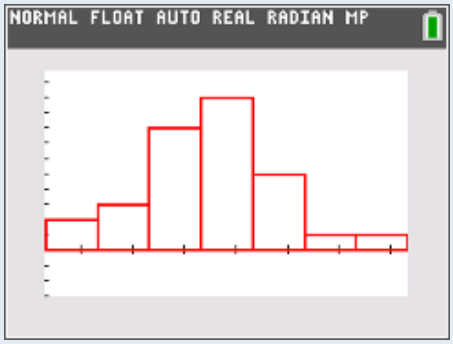
Using Ti 83/84 for Histograms

Below shows how the previous example can be completed using the ti-83/84 calculator.

- input data in to list L1: STAT → EDIT
- turn on statplot: 2nd → y= → ENTER → ENTER
- select histogram
- view graph: GRAPH → ZOOM → ZOOMSTAT







- change class intervals: WINDOW → enter values shown in picture below
- exam class intervals: TRACE → arrow left and right