

AP Students: Summer Assignments

In order for us to have solid math conversations on Statistics, Chapter 1 reviews many concepts that we have gone over in Algebra 1 or Algebra 2. Summer homework/preparation for all AP Stats students will **watch Chapter 1 Videos and complete the notes that accompany it. Then complete assignments based on videos and lesson. Please upload completed Assignments to Google Classroom.**

AP Statistics Chapter 1 Videos

AP Statistics Chapter 1 Note-Taking Guide

Chapter 1 Flashcards to Study

All summer assignments will be done in google classroom.

Google Classroom Code: 7wpxlsx

The videos are posted here and in google classroom.

Watch a video, take notes, complete assignment.

1.1A Worksheet

1.1B Worksheet

1.1C Worksheet

1.3A Worksheet

1.3B Worksheet

1.2A Worksheet

1.2B Worksheet

1.2C Worksheet

1.3C Worksheet

Chapter 1 Practice Test

If you have any questions, please email me sara.collison@matsuk12.us

1.1A Individuals are the _____ described by a set of data. Individuals may be _____.

1.1A Variable is any _____ of an individual. A variable can take different values for different individuals. There are _____ of variables based on the type of values that can be assigned to an individual:

A. **Categorical variables** places an individual into one of _____.
Gender is a categorical variable with the categories being male or female. A categorical variable can have a numerical value if it doesn't make sense to AVERAGE the value (i.e. zip codes, phone numbers)

B. **Quantitative variables** takes _____ for which it makes sense to find an average.

Sample problem:
Here is information about 10 randomly selected US residents from the 2000 census imported using Fathom software.

State	Number of Family Members	Age	Gender	Marital Status	Total Income	Phone Prefix
Kentucky	2	61	Female	Married	21000	949
Florida	6	27	Female	Married	21300	278
Wisconsin	2	27	Male	Married	30000	501
California	4	33	Female	Married	26000	232
Michigan	3	49	Female	Married	15100	877
Virginia	3	26	Female	Married	25000	154
Pennsylvania	4	44	Male	Married	43000	106
Virginia	4	22	Male	Never married	3000	475
California	1	30	Male	Never married	40000	158
New York	4	34	Female	Separated	30000	407

Problem:

(a) Who are the individuals in this data set?

(b) Identify each variable as categorical or quantitative. In what units were the quantitative variables measured?

(c) Describe the individual in the first row.

1.1A Distribution of a variable tells us _____ the variable takes and _____ it takes these values. We usually describe a distribution using a graph and numerical data.

1.1A **How to explore data**

- Examine each variable by _____
- Study the _____ among the variables
- Graph the distribution to look for _____
- Calculate _____ to look for patterns (centers, spread, etc)

1.1A **The four step process – guideline for free response questions**

State: what is the question you are _____? This allows us to find wording for what we are doing and later on what we have determined from our problems.

Plan: How will you go about answering the question? What _____ will you need to use and _____ you use them?

Do: Make _____ and carry out _____.

Conclude: Interpret what your calculations have told you about the _____. Always reference what this means in _____.

1.1A **Inference**

The whole purpose of statistics is to _____ that go beyond the data at hand. We are constantly looking at _____ about the world and people around us. That is what we call INFERENCE

1.1B **Association**

We say there is an association between two variables if _____ tend to occur in common with specific values of the other variable.

BE CAREFUL! Did you know that there is an association between cancer and attending school? Almost all people who have cancer went to school!

Working with CATEGORICAL DATA

1.1B **Tables** are a great way to display data _____. We will be working with two main types of tables:

A. Frequency table displays each category of the variable and the _____ of individuals that fall within that category

B. Relative frequency table displays each category of the variable and the _____ of the individuals that fall within that category.

Here is the distribution of bachelor's degrees awarded in 2007, according to

Major	Number of Majors
Business	327,531
Social sciences/History	164,183
Education	105,641
Health professions	101,810
Psychology	90,039
Visual and performing arts	85,186
Biological and biomedical sciences	75,151
Communication and related programs	74,783
Engineering	67,092
English language and literature	55,122
Other	377,554
Total	1,524,092

Major	Percent Of Majors
Business	21.5
Social sciences/History	10.8
Education	6.9
Health professions	6.7
Psychology	5.9
Visual and performing arts	5.6
Biological and biomedical sciences	4.9
Communication and related programs	4.9
Engineering	4.4
English language and literature	3.6
Other	24.8
Total	100.0

a.) the variable being measured is _____

b.) The first table is a _____ and the second table is a _____

1.1B Round off error
 When working with percentages, you will find that sometimes you end up with long decimals. When you round these decimals, you may find that the percentages _____.
 This is what we call round off error. They don't point to _____ in our work, just to the effects of rounding off results.

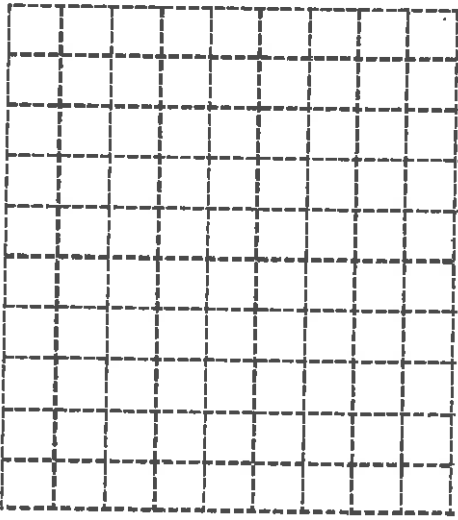
1.1B Percentages vs proportions
 When working with categorical data, we usually transform everything from frequency values to proportions (_____) this allows us to analyze how much of our whole group belong to certain categories. We can analyze these for patterns and even use this to predict certain events.

1.1B Bar graphs are a graphical display that can quickly illustrate the _____.
 Bar graphs can also be called bar charts and are used for _____. Usually the categories are listed on the x-axis and the frequency or relative frequency is listed on the y-axis.
 When making bar graphs do not forget to include:

- What variable is being measured
- What category does each bar represent
- Is it a frequency or a relative frequency – label the axis appropriately as a count, proportion, or percentage

Here are the percent of 15-18 year olds that own the following personal media devices, according to the Kaiser Family Foundation, make a well-labeled bar graph to display the data. Describe what you see.

Device	Percent who Own
Cell Phone	85%
MP3 Player	83%
Handheld Video Game Player	41%
Laptop	38%
Portable CD/Tape Player	20%



1.1B Pie charts
 Pie charts _____ be on the AP exam, but since they are widely used in the media, it is important to talk about them. Pie charts are usually not preferred by statisticians since they are hard to compare. Pie charts are not appropriate if we don't have all possible categories since the pie represents the "whole" picture. Pie charts are awkward to create unless you are using technology.
 Would it be appropriate to make a pie chart for the data above? Why or why not?

1.1B

What makes graphs good or bad?

Bad

- _____ on bars is not the same
- Using _____ instead of bars
- Leaving off the _____ of a scale
- "How to lie with statistics"

1.1C

Two-way tables

Two-way tables will represent the values of _____ variables. These are essential in calculating different types of _____. Like regular tables, these can be frequency or relative frequency tables. In two-way tables, it is okay to switch the locations of the variables.

1.1C

Marginal distributions

Marginal distributions isolate _____ variable and analyze the _____ of that one variable. To do this we analyze the frequency or the proportions in the total column for that variable.

1.1C

Sample: Super Powers

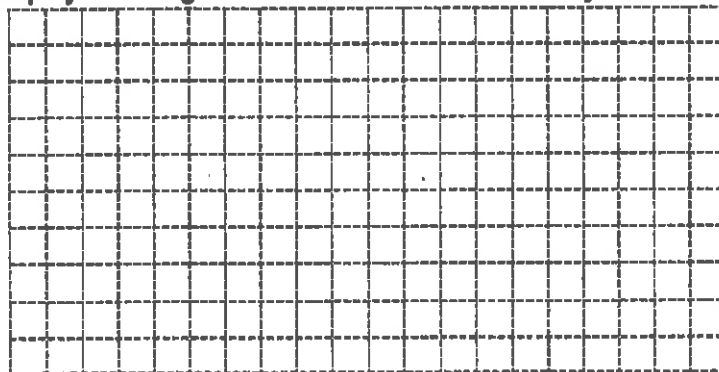
A sample of 200 children from the United Kingdom ages 9-17 was selected from the CensusAtSchool website (www.censusatschool.com). The gender of each student was recorded along with which super power they would most like to have: invisibility, super strength, telepathy (ability to read minds), ability to fly, or ability to freeze time.

Problem:

(a) Use the data in the two-way table to calculate the marginal distribution (as a proportion) of superpower preferences.

Responses	Female	Male	Total
Invisibility	17	13	30
Super Strength	3	17	20
Telepathy	39	5	44
Fly	36	18	54
Freeze Time	20	32	52
Total	115	85	200

(b) Make a graph to display the marginal distribution. Describe what you see.



1.1C

Conditional distributions

Conditional distributions follow closely to conditional _____ . With conditional distributions, we look at _____ for a variable, and analyze the distribution of the _____ in regards to that category. It allows us to compare distributions between categories.

Wording possibilities: Find the conditional distribution of _____ among (Total)

1.1C

Bar graphs for conditional distributions

To graphically compare conditional distributions, we use a _____ bar graph.

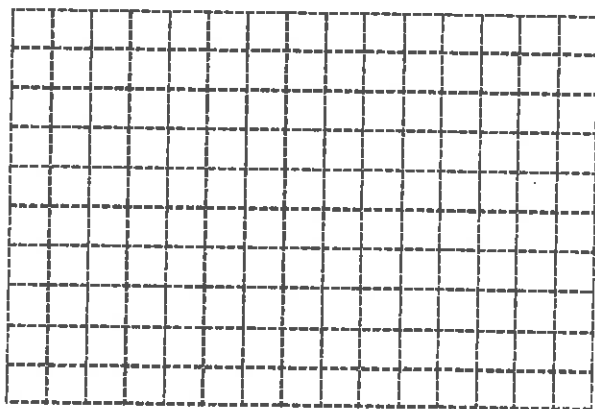
1.1C

Sample Continued: Super Powers

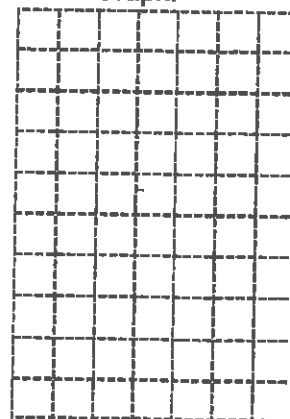
Calculate the conditional distribution of responses for each gender. Graph them as a side by side bar graph.

Responses	Female	Male	Total
Invisibility	17	13	30
Super Strength	3	17	20
Telepathy	39	5	44
Fly	36	18	54
Freeze Time	20	32	52
Total	115	85	200

Side by Side Bar graph:



Segmented Bar Graph:



Based on the survey data, can we conclude that boys and girls differ in their preference of superpower? Give appropriate evidence to support your answer.

1.3A

Measures of spread: VARIANCE (Var = s_x^2) and STANDARD DEVIATION (s_x)

When using _____ as a measure of center, it is appropriate to use variance and standard deviation as a measure of _____. Usually we only address standard deviation, but _____ is used to calculate standard deviation and has some important properties in later chapters.

The standard deviation measures the _____.

It is calculated by finding the **average of the squared distances (variance)** and then taking the square root.

$$s_x^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n-1}$$

$$s_x = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n-1}}$$

NOTE:

s_x refers to the standard deviation of a _____

σ refers to the standard deviation of a _____

s_x is always greater than or equal to _____ (only _____ when all values are the same number)

Standard deviation like mean is sensitive to _____

1.3A

Interpretation of STANDARD DEVIATION

"The _____ of the data values from the _____ of the distribution"

1.3A

Measures of spread: IQR

Since range only looks at the lowest and highest values (which may be outliers) we can improve our description of spread by looking at the _____.

1. The median divides your data into the first half (50%) and the second half.
2. Calculating the median of the _____ will divide the data into the first and second quarters (25%) of the data. This value is called the _____.
3. Calculating the median of the _____ will divide the data into the third and fourth quarters (25%) of the data. This value is called the _____.
4. The range between these values (3rd quartile - 1st quartile) is called the **Inter-quartile range**.

Note:

- The Inter-Quartile Range (IQR) is _____ to outliers since we cut out extreme values on both ends.
- The median itself is _____ included when calculating the quartiles

1.3A

Interpretation of IQR

"The middle _____ of the data has a range of _____ insert_IQR_____"

1.3A

Here is data for the amount of fat (in grams) for McDonald's chicken sandwiches:

Sandwich	Fat
Honey Mustard Snack Wrap® (Grilled)	9 g
Premium Grilled Chicken Classic Sandwich	10 g
McChicken ®	16 g
Honey Mustard Snack Wrap® (Crispy)	16 g
Premium Grilled Chicken Club Sandwich	17 g
Premium Crispy Chicken Classic Sandwich	20 g
Premium Crispy Chicken Ranch BLT Sandwich	23 g
Premium Crispy Chicken Club Sandwich	28 g

A.) Calculate and Interpret the IQR

B.) Calculate and Interpret the standard deviation

1.3A

Five number summary

The five number summary is a way to summarize a distribution _____ . It consists of the minimum value, the 1st quartile, the median, the 3rd quartile, and the maximum value.

Problem: What is the five number summary for the fat content of chicken sandwiches at McDonalds?

1.3A

OUTLIER RULE

How do we know for sure if a value is an outlier?

- Calculate the IQR
- Multiply that value by 1.5
- Lower Bound: subtract the 1.5IQR from the 1st quartile, anything lower than that value is an outlier
- Upper Bound: add the 1.5IQR from the 3rd quartile, anything above that value is an outlier

Problem: Determine if there is an outlier for chicken sandwiches fat content from McDonalds

1.3B

TECHNOLOGY CORNER Computing numerical summaries with technology

Let's find numerical summaries for the travel times of North Carolina and New York workers from the previous Technology Corner (page 61). We'll start by showing you the necessary calculator techniques and then look at output from computer software.

1. One-variable statistics on the calculator If you haven't done so already, enter the North Carolina data in L1/list1 and the New York data in L2/list2.

1. Find the summary statistics for the North Carolina travel times.

TI-83/84

- Press **2ND** **MODE** (CALC); choose 1:1-VarStats.
- Press **2ND** **ENTER**. Now press **2ND** **L1** (L1) and **ENTER**.

Press **ENTER** to see the rest of the one-variable statistics for North Carolina.

```
1-Var Stats
n=22
x̄=22.36363637
sx=16.919
sx²=285.845
σx=13.71446706
n=22
```

```
1-Var Stats
n=22
x̄=22.36363637
sx=16.919
sx²=285.845
σx=13.71446706
n=22
```

TI-89

- Press **2ND** (Calc); choose 1:1-Var Stats.
- Type list1 in the list box. Press **ENTER**.

2. Repeat Step 1 using L2/list2 to find the summary statistics for the New York travel times.

```
1-Var Stats
n=20
x̄=16.25
sx=11.5
sx²=132.25
σx=9.126194673
n=20
```

```
1-Var Stats
n=20
x̄=16.25
sx=11.5
sx²=132.25
σx=9.126194673
n=20
```

1.3B

Which gender is taller, males or females?

A sample of 14-year-olds from the United Kingdom was randomly selected. Here are the heights of the students (in cm):

Male: 154, 157, 87, 163, 167, 159, 169, 162, 176, 177, 151, 175, 174, 165, 165, 183, 180

Female: 160, 169, 152, 167, 164, 163, 160, 163, 169, 157, 158, 153, 161, 165, 165, 159, 168, 153, 166, 158, 158, 166

Calculate the numerical summaries for each gender:

	n	\bar{x}	S_x	Min	Q1	M	Q3	Max
Males								
Females								

Determine if there are any outliers:

1.2A

Dotplots

- Make a number line
- Plot the correct number of dots above the value
- Analyze the distribution
- Best used for _____ sets of data

Here is data regarding the number of family members for 10 randomly selected US residents from the 2000 census: 2,6,2,4,3,3,4,4,1,4. Construct a dotplot of the data:



1.2A

Stemplots

- Best used for _____ sets of data
- Each value is separated into a stem (_____) and a leaf (_____)
- Write stems in a vertical column with the smallest at the top
- Draw a vertical line to the right of the column
- DO NOT _____ any stems, even if there is no data value for that particular stem.
- Write each leaf in the row to the right of the stem in _____ order out to the right
- Provide a _____ that explains how to read what each value represents

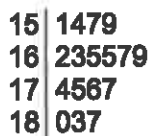
Sample: Which gender is taller, males or females?

A sample of 14-year-olds from the United Kingdom was randomly selected. Here are the heights of the students (in cm):

Male: 154,157,87,163,167,159,169,162,176,177,151,175,174,165,165,183,180

Female: 160,169,152,167,164,163,160,163,169,157,158,153,161,165,165,159,168,153,166,158,158,166

Here is a stemplot for Male heights



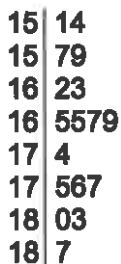
Key: 15 1 represents a male who is 151 cm tall.

1.2A

Special Case: Splitting stems

- If you have less than _____ stems OR if one or more stems has _____ leaves – split stems
- Stem is used twice – the first one will have leaves _____ and the second will have leaves _____

Here is a stemplot for Male heights with split stems:



Key: 15 1 represents a male who is 151 cm tall.

1.2A

Special Case: Back to back stemplots

- Allows you to break apart your data into two categories (like male and female) and graph one category on one side of the stem and the other category on the other side of the stem.
- This allows you to compare the distribution for the two categories.

Here is a back-to-back stemplot comparing male and female heights:

Female		Male
332	15	14
98887	15	79
433100	16	23
99876655	16	5579
	17	4
	17	567
	18	03
	18	7

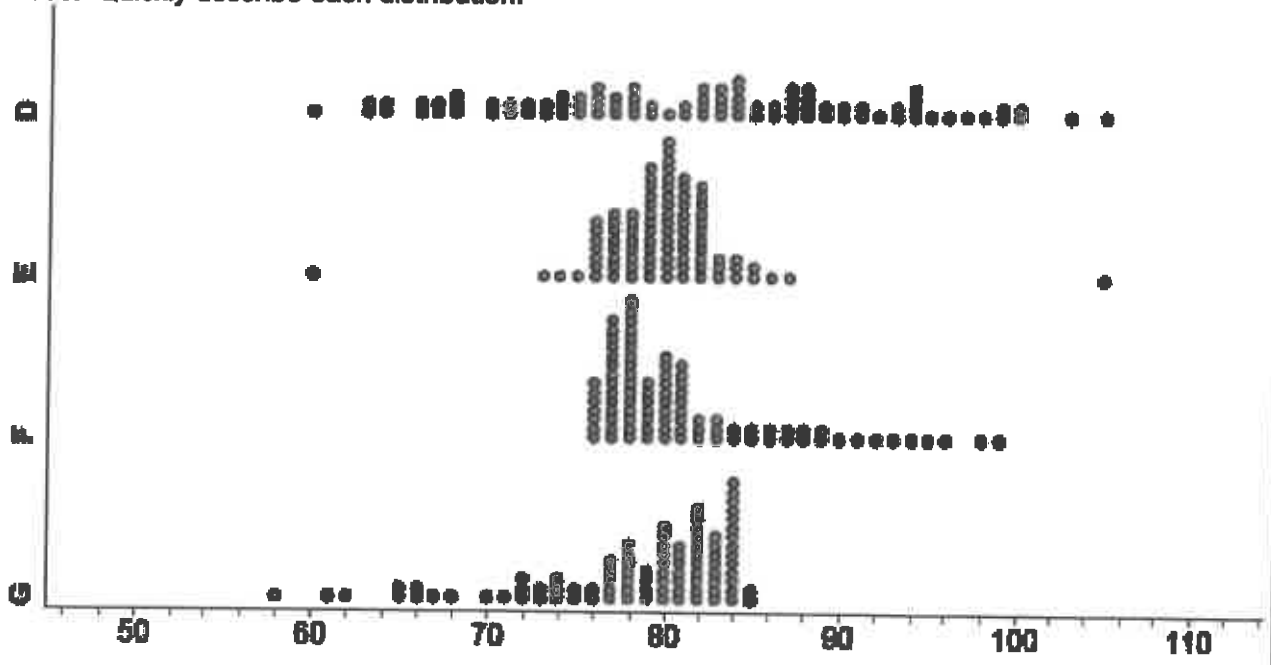
Key: 15|1
represents a student who is 151

1.2A

How to examine a distribution

- **Shape:** Unimodal (one main peak) can be symmetric (peak in middle), or skewed right (peak to left, tail to right) or skewed left (peak to right, tail to left), bimodal, multimodal, or uniform (all values are roughly the same).
- **Center:** Mode (peak), Median (50% of values above and below, can look at as a midpoint), and mean (average, balancing point)
- **Spread:** Range (highest - lowest), Interquartile range, and Standard Deviation. Also referred to as _____
- **Outliers:** Values outside the main group of values or extreme values

Sample: Brian and Jessica have decided to move and are considering seven different cities. The dotplots for four cities show the daily high temperatures for June, July, and August for each of these cities. Quickly describe each distribution.



1.2B

Histograms

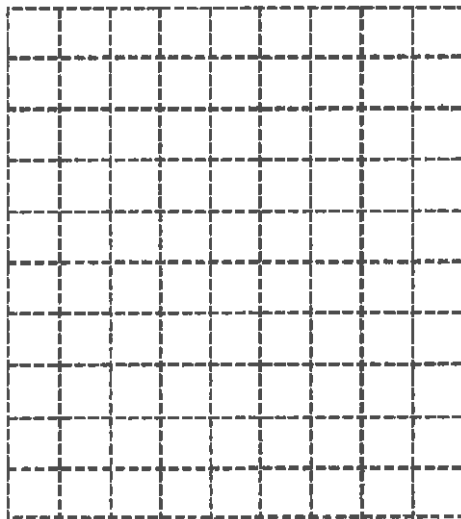
- Histograms are best used for _____ sets of data, and data that does not have _____ values
- Histograms take values and places them in "bins", which is an _____.
- Each bar represents the _____ in that interval or bin.
- All intervals or bins should be _____ in width
- The lower boundary for the bin is always _____ in the interval, but the second boundary is _____. It will be the beginning for the next bin.

Sample: Here are data for the amount of fat (In grams) for McDonald's beef sandwiches:

Sandwich	Fat (g)
Double Quarter Pounder® with Cheese	42 g
Angus Mushroom & Swiss	40 g
Angus Bacon & Cheese	39 g
Angus Deluxe	39 g
Big Mac®	29 g
Big N' Tasty® with Cheese	28 g
Quarter Pounder® with Cheese	26 g
McRib ®	26 g
Big N' Tasty®	24 g
Double Cheeseburger	23 g
McDouble	19 g
Quarter Pounder®	19 g
Mac Snack Wrap	19 g
Cheeseburger	12 g
Hamburger	9 g

Create a histogram with bin widths of 5

Describe the distribution.



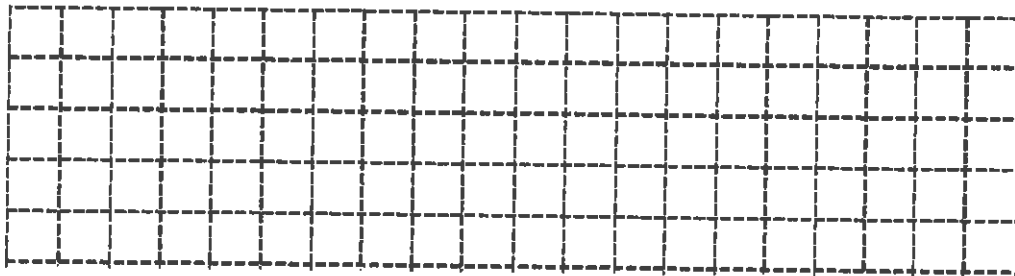
1.2B

Boxplot – modified

A boxplot is a graphical representation of the _____. A modified boxplot will also show if and where there are _____.

Problem: Create boxplots for the number of contacts for male and female high school students.

	<i>n</i>	\bar{x}	<i>s_x</i>	min	<i>Q1</i>	<i>M</i>	<i>Q3</i>	max
Male	20	122.15	73.83	27	64.5	109.5	167.5	290
Female	14	126.71	65.75	22	83	129	180	218



Describe the distributions:

1.2C

Histogram on the calculator

TECHNOLOGY CORNER Histograms on the calculator

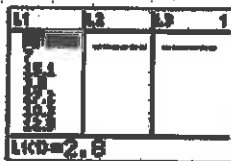
TI-83/84

TI-89

1. Enter the data for the percent of state residents born outside the United States in your Statistics/List Editor.

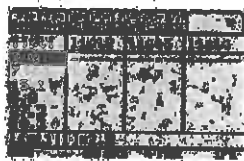
• Press **[2ND]** and choose 1:edit...

• Type the values into list L1.



• Press **[2ND]** and select stats/List editor.

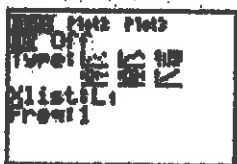
• Type the values into list1.



2. Set up a histogram in the Statistics Plots menu.

• Press **[2ND]** **[F2]** (STAT PLOT).

• Press **[ENTER]** or **[↓]** to go into Plot1.



• Press **[F2]** and choose 1:Plot setup...

• With Plot1 highlighted, press **[F1]** to define.



Set Hist. Bucket Width to 5.

What percent of your home state's residents were born outside the United States?

The percentage for each state is listed below:

2.8,7.0,15.1,3.8,27.2,10.3,
12.9,8.1,18.9,9.2,16.3,5.6,
13.8,4.2,3.8,6.3,2.7,2.9,3.2,
12.2,14.1,5.9,6.6,1.8,3.3,
1.9,5.6,19.1,5.4,20.1,10.1,
21.6,6.9,2.1,3.6,4.9,9.7,5.1,
12.6,4.1,2.2,3.9,15.9,8.3,
3.9,10.1,12.4,1.2,4.4,2.7

3. Use ZoomStat (ZoomData on the TI-89) to let the calculator choose classes and make a histogram.

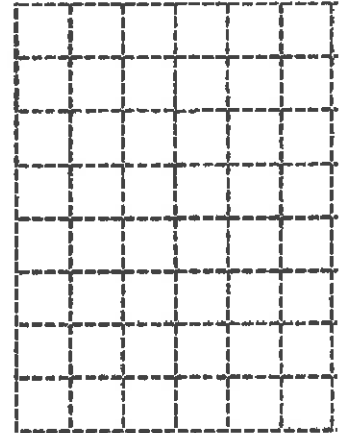
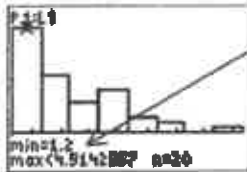
• Press **[ZOOM]** and choose 9:ZoomStat.

• Press **[TRACE]** and **[◀▶]** to examine the classes.

• Press **[F5]** (ZoomData).

• Press **[F1]** (Trace) and **[◀▶]** to examine the classes.

Note the calculator's unusual choice of classes.



4. Adjust the classes to match those in Figure 1.16, and then graph the histogram.

• Press **[WINDOW]** and enter the values shown.

• Press **[GRAPH]**

• Press **[TRACE]** and **[◀▶]** to examine the classes.

• Press **[◀▶]** **[F2]** (WINDOW) and enter the values shown.

• Press **[◀▶]** **[F2]** (GRAPH)

• Press **[F1]** (trace and **[◀▶]** to examine the classes)

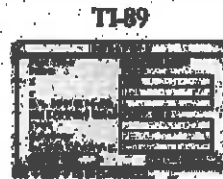
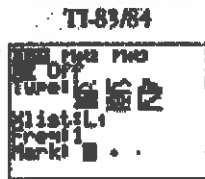


1.2C

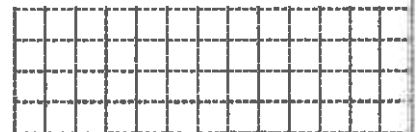
TECHNOLOGY CORNER Making calculator boxplots

The TI-83/84 and TI-89 can plot up to three boxplots in the same viewing window. Let's use the calculator to make side-by-side boxplots of the travel time to work data for the samples from North Carolina and New York.

1. Enter the travel time data for North Carolina in L1/list1 and for New York in L2/list2.
2. Set up two statistics plots: Plot1 to show a boxplot of the North Carolina data and Plot2 to show a boxplot of the New York data.



Travel times for a random sample for 15 adults in North Carolina in minutes:
5, 10,10,10,10,12,15,20,20,25,
30,30,40,40,60

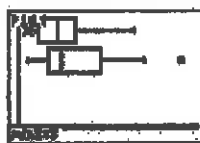


Note: The calculator offers two types of boxplots: a "modified" boxplot that shows outliers and a standard boxplot that doesn't. We'll always use the modified boxplot.

3. Use the calculator's Zoom feature to display the side-by-side boxplots. Then Trace to view the five-number summary.

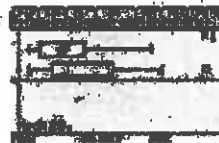
TI-83/84

- Press **[ZOOM]** and select 9 : ZoomStat.
- Press **[TRACE]**



TI-89

- Press **[F5]** (ZoomData).
- Press **[F1]** (Trace).



1.3C

Comparing mean and median

- In _____ distributions, the mean and median are roughly the same
- In skewed distributions, the mean is pulled towards the _____

Identify which value is the mean and which value is the median:



- When distributions are skewed right, the mean will be _____ than the median
- When distributions are skewed left, the mean will be _____ than the median
- When distributions are symmetric, the mean will be _____ to the median

1.3C

What is the best measure of center and spread?

1. Is your data symmetric OR skewed?
2. Are there outliers?

1.3C

Problem: Who Has More Contacts—Males or Females?

The following are five number summaries for the number of contacts that a sample of high school students had in their cell phones. Describe what these values are telling you.

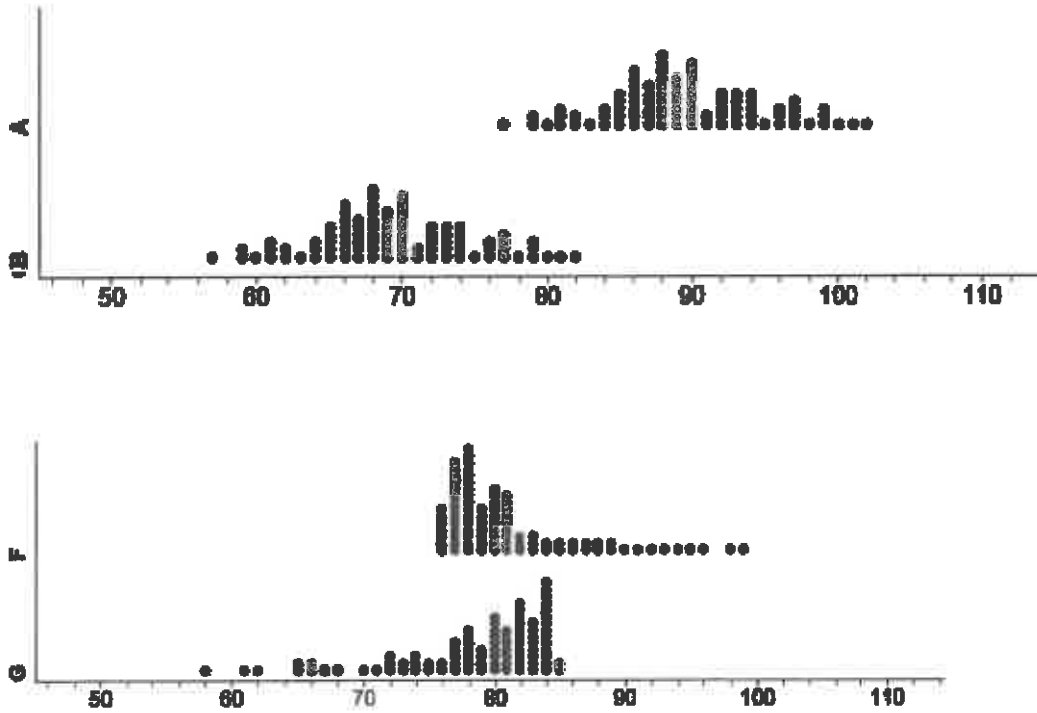
	n	\bar{x}	s_x	min	$Q1$	M	$Q3$	max
Male	20	122.15	73.83	27	64.5	109.5	167.5	290
Female	14	126.71	65.75	22	83	129	180	218

- A.) Are there any outliers in either distribution?
- B.) Based on the numerical summary, what shape is the distribution of males?
- C.) Based on the numerical summary, what shape is the distribution of females?
- D.) Compare the distributions.

1.3C

Comparing distributions

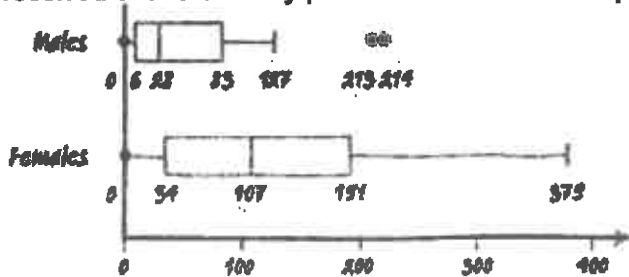
Address all four characteristics and use comparison phrasing like "about the same as" or "much greater than". Compare the distributions for the cities:



1.3C

Comparing boxplots

A group of students investigated their belief that females text more than males. They asked a random sample of students from their school to record the number of text messages sent and received over a two day period. Here are the boxplots:



Compare the distributions:

1.3C

Comparing stemplots

Female	Male
332	15 14
98887	15 79
433100	16 23
99876655	16 5579
	17 4
	17 567
	18 03
	18 7

Key: 15|1
represents a
student who is 151
cm tall.