ANNE ARUNDEL COUNTY PUBLIC SCHOOLS
ELEVATING ALL STUDENTS... ELIMINATING ALL GAPS

Future AP Statistics Student,

Congratulations on making an excellent decision to study Statistics! The science of statistics is relevant in so many areas of your world. Statistics is used in sports, business, political science, research, and data analysis. Throughout the year we will explore statistics through a culturally relevant lens.

Advance Placement Statistics is a demanding course and the Statistics teaching team at South River has high expectations for all students. We want all of our students to strive for a five on the AP exam. In order to obtain this goal the AP Statistics journey must start this summer. We have prepared summer lessons that covers the content in the first 2 chapters of the text. The summer material is a review of information learned in Common Core Algebra II.

The AP Statistics classes are conducted similar to college courses. We expect you to read and take notes BEFORE the topic is covered in class. This technique will be used throughout the entire year. In the packet provided you will see note outlines and practice problems for each section in chapter 1 and chapter 2. Please use the answers in the back of the book to check your work. Please complete this work by the first week of school. You will be assessed on these chapters within the first 2 weeks.

You also need to join the South River AP Statistics google classroom. Helpful videos will be posted to guide you through the material. This will also be a place to develop a class forum. If you do not have a google classroom account you can use your school email address and create a student account at classroom.google.com. Once you have created an account, join the AP Statistics group using the code **cdk20p9**

Lastly, you will need to obtain a book from Mrs. Smith in room C270. This can be done at the beginning of NEST any day or during the 2nd block of NEST on Tuesday and Friday**.** If you have any questions please contact Mrs. Smith at mrsmith2@aacps.org.

With Excitement,

The AP Statistics Teaching Team

Name:_____ Date:_____

Section Title:_____ p. 1-6

**LEARNING OBJECTIVES: I will be able to....**

Individuals:

Variables:

| Categorical | Quantitative |
|---|---|
| Definition: | Definition: |
| Examples: | Examples: |

Distribution:

How to Explore Data:

Other:

**Reflection:**

Answer the "Check your Understanding" on p. 4-5?

Name:_____ Date:_____

Section Title:_____ p. 7-20

**LEARNING OBJECTIVES: I will be able to....**

Ways to Display Categorical Data:


ALWAYS remember to LABEL your Graphs. Forgetting to label graphs will cause a deduction in your score!

| Frequency Tables | Bar Graphs |
|---|---|
| | |
| **Pie Charts** | **Two Way tables** |
| | |

Difference between Marginal and Conditional Distribution

Relationships between Categorical Variables:

Other:

**Reflection:**

Answer the "Check your Understanding" on p.14 and p. 18?

Name: _____ Date:_____

Section Title: _____ p. 25-40

**LEARNING OBJECTIVES: I will be able to....**


ALWAYS remember to LABEL your Graphs. Forgetting to label graphs will cause a deduction in your score!

How to Make a Dot Plot:

Examining the Distribution of Quantitative Variables:

| | |
|---|---|
| Shape | |
| Center | |
| Spread | |
| Outliers | |


When asked on the AP exam to describe a distribution you must remember to address all 4 of the items above in order to get full credit on a question. The acronym SOCS might help. If comparing two distributions, make sure to compare all 4 characteristics!

Stemplots:

Histogram: (histograms will be covered in depth during the next class)

Steps to Creating a Histogram:

**Reflection:**

Answer the "Check your Understanding" on p.29 and p. 32-33?

Name: _____ Date:_____

Section Title: _____ p. 48-67

**LEARNING OBJECTIVES: I will be able to....**


FORMULAS are provided on the formula sheet! Know what is on the formula sheet!

## Measures of Center:

When to choose the mean or the median?

## Measures of Spread:

IQR-

Calculating Outliers:

**The five number summary and Box Plots:**

**Standard Deviation:**

Choosing measures of center and spread?

**Reflection:**

Answer the "Check your Understanding" on p.53 and p. 59?

Name:_____ Date:_____

**LEARNING OBJECTIVES: I will be able to....**

Section Title:_____ p. 83-99

**Measuring Position: Percentile**


WORDS ARE IMPORTANT: Percentiles are a specific location in a distribution, so a data point isn't "in" the $84^{th}$ percentile. Rather, it is "at" the $84^{th}$ percentile

Cumulative Frequency Graphs:

Connection between Percentiles and Quartiles:

**Measuring Position: z-scores**

**Transforming Data:**

| | |
|---|---|
| **Effect of adding (or subtracting) a constant:** | |
| **Effect of Multiplying (or dividing) by a constant:** | |

**Reflection:**

Answer the "Check your Understanding" on p.89 and p. 91?

Name:_____ Date:_____

Section Title:_____ p. 103-127

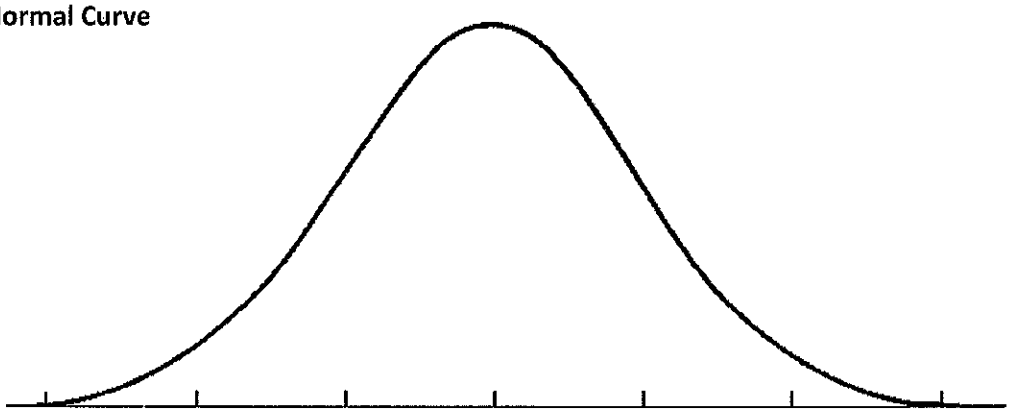
**LEARNING OBJECTIVES: I will be able to....**

**Density Curves**

Mean vs Median on a Density Curve

CYU p. 107

**Normal Distribution and Normal Curve**




The empirical rule (68-95-99.7) ONLY apples to Normal curve. Do not try to apply it to a non-normal distribution. Chebyshev's inequality (mentioned on p. 112) is NOT on the AP exam.

Complete CYU p. 112

**Standard Normal Distribution**


Always sketch the area under the curve. This will help you avoid making mistakes like reporting the area to the left of a boundary point when asked about the area to the right.

CYU p. 121

# Chapter 1: Exploring Data

## 1.1 Describing Distributions with Graphs

Statistics is the science of data. We begin our study of statistics by mastering the art of examining data. In this chapter of YMS, you learn how to make a number of displays including dotplots, stemplots, histograms, and ogives. It is good practice to construct these plots by hand to gain a better understanding of their meaning and connections to your data. However, once you've mastered construction by hand, our TI's are capable of making basic univariate plots such as histograms, boxplots, and modified boxplots. *Note: Do not rely on your calculator to make these plots until you have mastered constructing them by hand!*

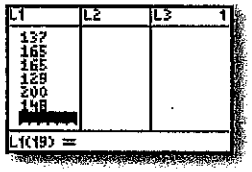
### Displaying Univariate Data

Consider the following data on the Survey of Study Habits and Attitudes (SSHA) scores for 18 female college students. The test evaluates motivation, study habits, and attitudes toward school:

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 154 | 109 | 137 | 115 | 152 | 140 | 154 | 178 | 101 |
| 103 | 126 | 126 | 137 | 165 | 165 | 129 | 200 | 148 |

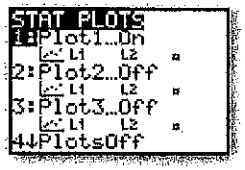
To make a histogram or boxplot on our TI, we must first enter our data. Entering data on the TI is easy. Data can be stored in Lists in a spreadsheet program under the [STAT] menu.

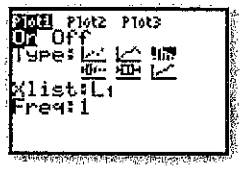
1. Press [STAT] **1: Edit...**
2. Enter the "SSHA" data into **L1**
3. Enter all 18 values into the list, pressing [ENTER] after each value.




To view a plot of the data, you need to set up your statistics plots:

1. Press , **Y= (STAT PLOT)**
2. Select **1: Plot1...**
3. Press [ENTER] to highlight **On**
4. Select the **Histogram** option under **Type:**



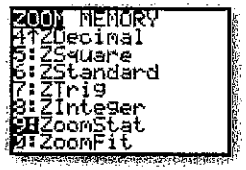



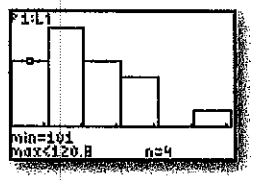



You are now ready to view a histogram of the "SSHA" data. To do this, you need to set your window to the appropriate values. You can do this by changing the parameters in the **WINDOW** mode, or you can **Zoom** directly to the data.

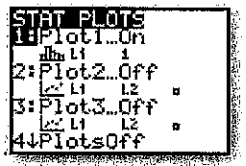
**To zoom directly to the histogram:**

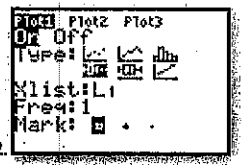
1. Press **ZOOM**
2. Select **9: ZoomStat**
3. Describe the plot *in the context of the problem*
4. Press **TRACE** to see categories and frequencies



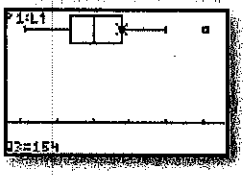

Using the "SSHA" data, select **modified boxplot** to get another view of the distribution.

1.



2.


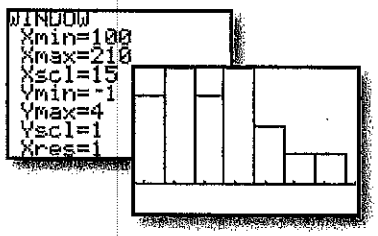

3.



4.



**To set the window parameters for a histogram or boxplot yourself:**

1. Press **WINDOW**
2. Set **Xmin** and **Xmax** to reflect the minimum and maximum of your dataset
3. Set **Ymin** to −1 and **Ymax** to the largest frequency
4. Set **Yscl** to equal your desired category width
5. Press **Graph** to see your plot



## Comparing Data Displays

Throughout the course of your studies, you may be asked to compare sets of univiarate data. Your calculator has the ability to display two boxplots on the same screen to allow for easy comparison. *Again, do not rely on your calculator until you understand how to do it by hand!*

Consider the following data on home run counts for Barry Bonds and Hank Aaron.

**Barry Bonds**

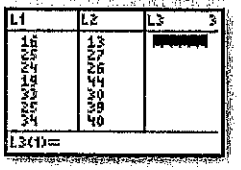
| 16 | 25 | 24 | 19 | 33 | 25 | 34 | 46 | 37 | 33 | 42 | 40 | 37 | 34 | 49 | 73 | 46 | 45 | 45 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|

**Hank Aaron**

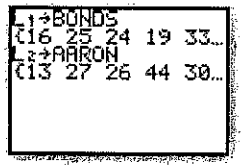
| 13 | 27 | 26 | 44 | 30 | 39 | 40 | 34 | 45 | 44 | 24 | 32 | 44 | 39 | 29 | 44 | 38 | 47 | 34 | 40 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|

1. Enter the Bonds data into **L1**
2. Enter the Aaron data into **L2**

   You do not need to put the data in order. If this is desired, you can "sort" the list using the **SortA(** command in the **LIST** menu.

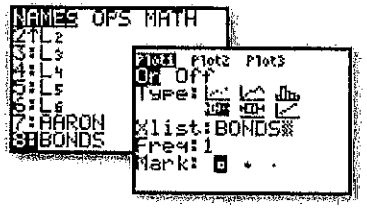


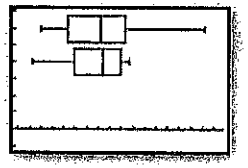

3. Save the data for future reference
4. On your "homescreen", enter the following
   **2ND** [1] **(L1)** [STO▸] **ALPHA B O N D S** [ENTER]
   **2ND** [2] **(L2)** [STO▸] **ALPHA A A R O N** [ENTER]
   Your lists have now been stored for future reference.




5. Press **2ND Y= (STAT PLOT)**
6. Set **Plot1 On Type: Modified Boxplot**
7. Set **Xlist: to 2ND STAT (LIST) X:BONDS**

8. Press **2ND Y= (STAT PLOT)**
9. Set **Plot2 On Type: Modified Boxplot**
10. Set **Xlist: to 2ND STAT (LIST) X:AARON**



11. Press **ZOOM 9:ZoomStat**

12. Compare the home run counts for Bonds (top) and Aaron (bottom). Don't forget to interpret the *SOCS* (Shape, Outliers, Center, Spread) for each batter!



# 1.2 Describing Distributions with Numbers

When you first encounter a dataset, it is a good habit to study a graphical display and estimate the SOCS. However, for a more detailed understanding of data, we must calculate numeric summaries of the center and spread. *Note: Be sure you understand how the following measures are calculated before relying on the TI to do the mechanics for you.*

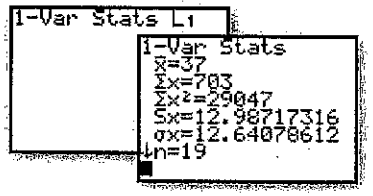
The most common measures of center for a dataset are **mean** ($\bar{x}$) and **median** (Q2). The most common measures of spread/variability for a dataset are **range** (max-min), **interquartile range "IQR"** (Q3-Q1), and **standard deviation** ($s_x$).

## Calculating Numeric Summaries

The calculation of each of these measures, especially the standard deviation, can be quite tedious. Thankfully, the TI can automate those calculations for us. Like plotting data, the calculator requires that you enter the dataset before it can report a numeric summary. If you haven't done so already, enter the Bonds and Aaron data into [STAT] **Edit… L1** and **L2**, respectively.

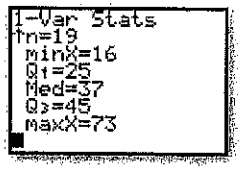
1. Enter data in to into [STAT] **1:Edit…**
2. Press [STAT] **CALC 1:1-Var Stats** [ENTER]


3. Your homescreen should read **"1-Var Stats"**
4. Press **2ND** [1] **(L1)** [ENTER]
5. A numeric summary of the Bonds data should appear.
6. Repeat Steps 2 through 4 for **L2** to get a numeric summary of the Aaron data.




7. Scroll down on each numeric summary to see the 5-number summary.

*Remember to interpret the numeric summary in the context of the problem!*



# Chapter 2: Statistical Models for Distributions

## 2.2 Normal Distributions

In Chapter 2 of YMS, we learn that distributions of data can be approximated by a mathematical model known as a **density curve**. In this section, we learn that many sets of data, probability applications, and sampling situations can be modeled by normal distributions. If a set of data is *approximately normal*, we can use *standardized normal calculations* to determine relative standing, percentiles, and probabilities. Like our data displays and numeric summaries, it is important that you understand how to perform the following calculations by hand before relying on the calculator.

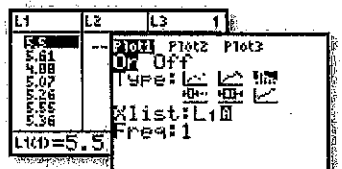
### Approximating a Normal Curve

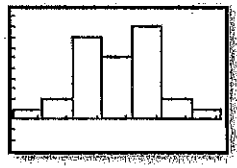
Consider the following measurements of the density of the earth taken by Henry Cavendish in 1798. Use your calculator to construct a histogram of these data.

**Density of the earth (as a multiple of the density of water)**

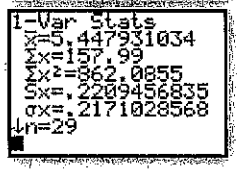
| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 5.5 | 5.61 | 4.88 | 5.07 | 5.26 | 5.55 | 5.36 | 5.29 | 5.58 | 5.65 |
| 5.57 | 5.53 | 5.62 | 5.29 | 5.44 | 5.34 | 5.79 | 5.10 | 5.27 | 5.39 |
| 5.42 | 5.47 | 5.63 | 5.34 | 5.46 | 5.30 | 5.75 | 5.68 | 5.85 | |

1. Enter the Density data into **L1**
2. Set up **STAT PLOT** to view a histogram
3. Use **ZOOM 9:ZoomStat** to view the distribution
4. Calculate the mean and standard deviation by using [STAT] **Calc 1:1-Var Stats**



Note the shape of this distribution. It appears it can be approximated by a **normal distribution** with mean = 5.45 and standard deviation = 0.22.



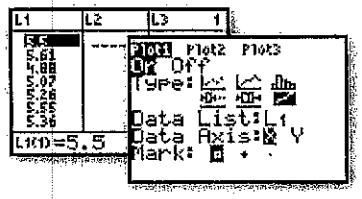

Remember, just because data appears normally distributed, that doesn't make it so. We can assess the normality of data by constructing a different stat plot on our calculator. This plot, known as a **normal quantile plot** or **normal probability plot** can be used to assess the adequacy of a normal model for a data set. If the plot appears to have a linear pattern, a normal model is reasonable. If the plot appears nonlinear, we believe the data may be from a nonnormal distribution.

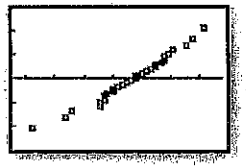


### Assessing Normality

Before performing normal calculations, let's check the normality of our Density data. If you haven't done so already, enter the Cavendish Density data into **L1**.

1. Enter Density data into **L1**
2. Set up **STAT PLOT** to construct a normal probability plot of the data in **L1**. This is the last option under **Type:**
3. Use **ZOOM 9:ZoomStat** to see the plot
4. Assess the normality of the Density data



Since this plot appears approximately linear, we can conclude that it is reasonable to assume the data are from a normal distribution.

### Normal Calculations

The **Empircal (68-95-99.7) Rule** for approximately normal distributions is a handy tool for calculating relative standing, percentiles, and probabilities. However, many problems we encounter will involve observations that don't fall exactly 1, 2, or 3 standard deviations away from the mean. In these cases, we'll need to perform a standard normal calculation.

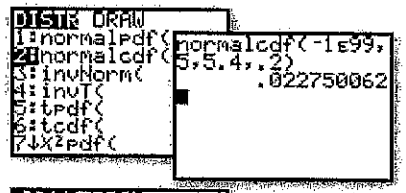
Standard normal calculations are fairly easy to perform by hand, but our calculator does have the ability to automate the process. Again, please be sure to practice these calculations by hand before using the calculator. In some cases, calculating standardized scores and looking up values on a table may be quicker than using the calculator!

Standard normal calculations are performed in the **Distribution** menu of the calculator. You will want to use the "normal cumulative density function". To use this, you must enter the following parameters in order **normalcdf(min, max, mean, standard deviation)**. If your min is not defined to the left, use **-1E99**. If your max is not defined to the right, use **1E99**. The mean and standard deviation should be the given or calculated mean and standard deviation of your approximately normal distribution.

Let's use the Density data to answer some questions about the proportion of observations that would satisfy certain conditions. We have already determined this data is approximately normally distributed with a mean=5.4 and standard deviation of .2.

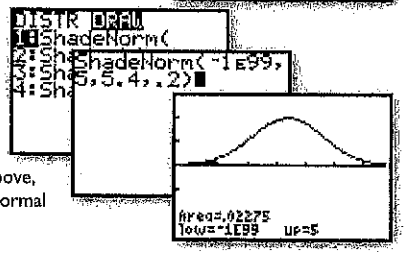
What proportion of density observations fall below 5?

1. Press **2ND** [VARS] **(DISTR)**
2. Choose **2:normalcdf(**
3. Enter **normalcdf(-1E99, 5, 5.4, .2)**




If you would like to see the curve:

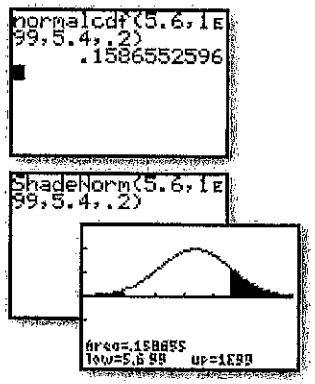
1. **2ND** [VARS] **(DISTR)**
2. Choose **DRAW 1:ShadeNorm(**
3. Enter **ShadeNorm(-1E99, 5, 5.4, .2)**



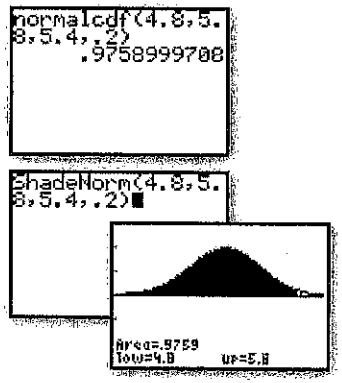

Notice in both cases we find approximately 2.3% of observations will fall below a measurement of 5.

This process can be used to calculate the proportion of observations falling above, below, or between any two points in a normal distribution.

Find the proportion of density observations above 5.6:




Find the proportion of density observations between 4.8 and 5.8:





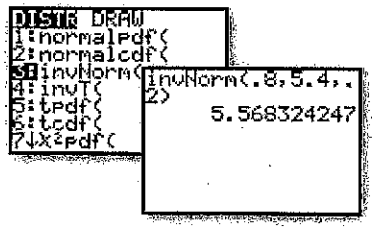

## Inverse Normal Calculations

A number of times throughout the course, you will be asked to determine what point cuts off a desired proportion in a normal distribution. For example, we may want to know what value determines the 80th percentile in our density data. That is, what value falls at a point such that 80% of observations fall below it?

To answer this question on our calculator, we need to use the "inverse normal" function with the following parameters **invNorm(left area, mean, standard deviation)**.

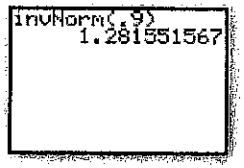
1. Press **2ND VARS (DISTR)**
2. Select **3:invNorm(**
3. Enter **invNorm(.8, 5.4, .2)**

We can see that 80% of density observations will fall below 5.57.




## Finding z-scores with invNorm

If you would like to find a z-score corresponding to a desired percentile, you can simply enter **invNorm(percentile)**. If you don't enter a mean or standard deviation in the **invNorm, normcdf(, or ShadeNorm(** functions, the calculator assumes you are referring to a standard normal distribution with mean = 0 and standard deviation = 1.




For example, the z-score that determines the 90th percentile in a standard normal distribution is **invNorm(.9) = 1.281551567**