AP® Statistics:
TI83+/84+ User Guide

For use with "The Practice of Statistics" by Yates, Moore, Starnes

Jason M. Molesky
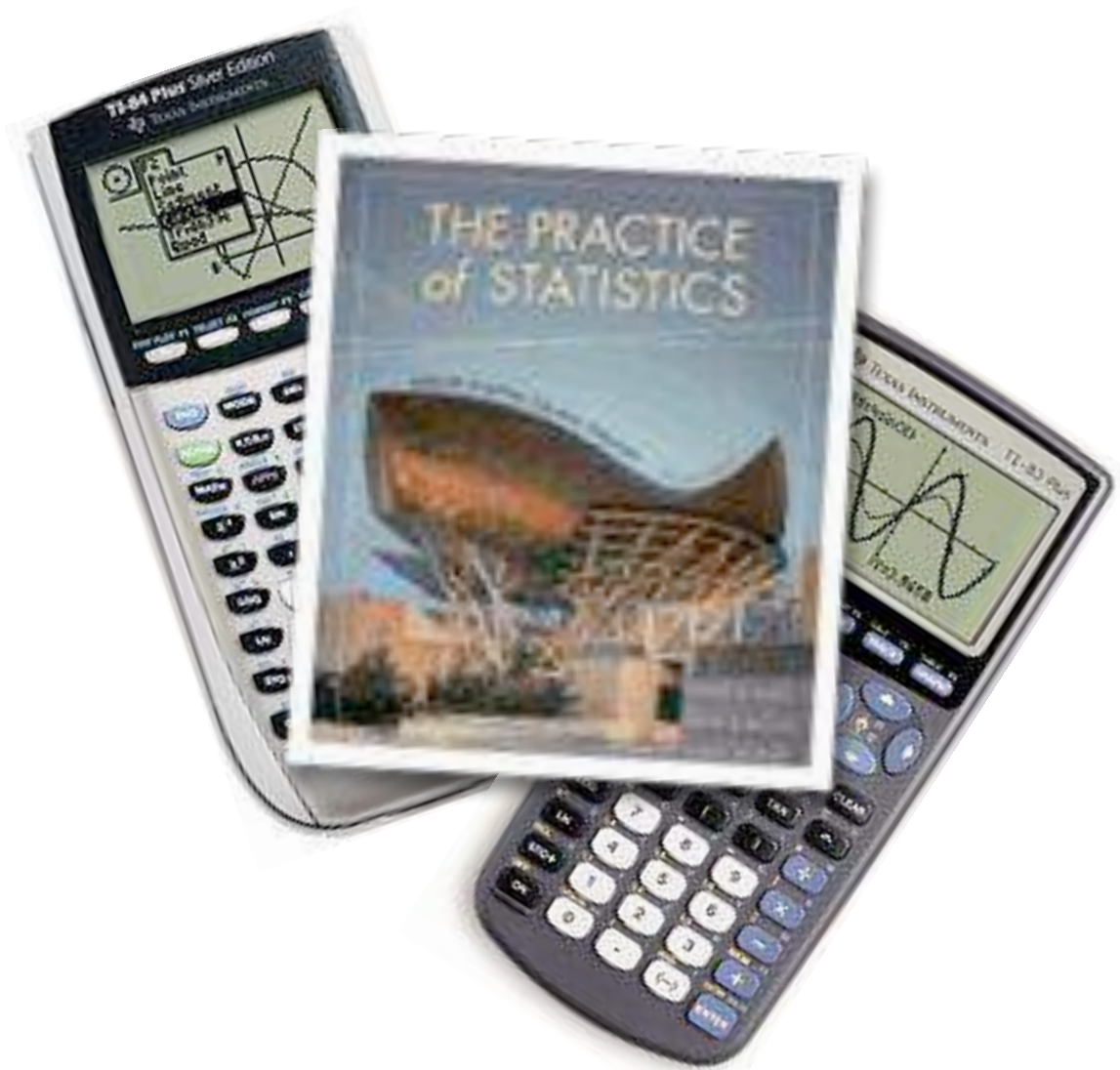**Lakeville South High School • Lakeville, MN • web.mac.com/statsmonkey**

Table of Contents

Note:

Most of the examples that follow are taken directly from examples and problems found in your textbook, "*The Practice of Statistics*" by Yates, Moore, and Starnes. You will recognize most of the datasets from your reading and homework. The purpose of this guide is to provide you, the student, with an abbreviated set of examples to assist you in using your TI graphing calculator in your study of AP Statistics.

Keep this guide in an accessible place as it will be a helpful reference tool as you prepare for the Advanced Placement Exam. Should you need further explanation of any of the topics, be sure to refer to the Technology Toolboxes in your textbook or see your teacher for help.

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!_
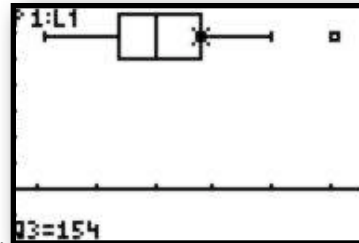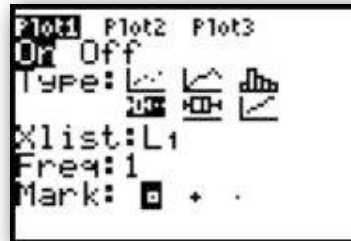
Displaying Univariate Data

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

```
ZOOM MEMORY
4↑ZDecimal
5:ZSquare
6:ZStandard
7:ZTrig
8:ZInteger
9:ZoomStat
0:ZoomFit
```

```
P1:L1

min=101
max<120.8         n=4
```
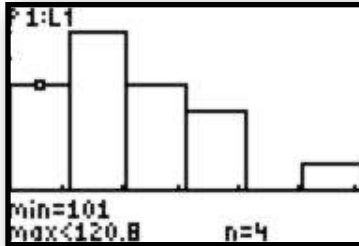
Using the "SSHA" data, select **modified boxplot** to get another view of the distribution.

```
STAT PLOTS
1:Plot1…On
   L1      1
2:Plot2…Off
   L1   L2
3:Plot3…Off
   L1   L2
4↓PlotsOff
```
1.

```
Plot1  Plot2  Plot3
On  Off
Type:
Xlist:L1
Freq:1
Mark: ■  +  ·
```
2.

```
ZOOM MEMORY
4↑ZDecimal
5:ZSquare
6:ZStandard
7:ZTrig
8:ZInteger
9:ZoomStat
0:ZoomFit
```
3.

```
P1:L1

Q3=154
```
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

```
WINDOW
Xmin=100
Xmax=210
Xscl=15
Ymin=-1
Ymax=4
Yscl=1
Xres=1
```
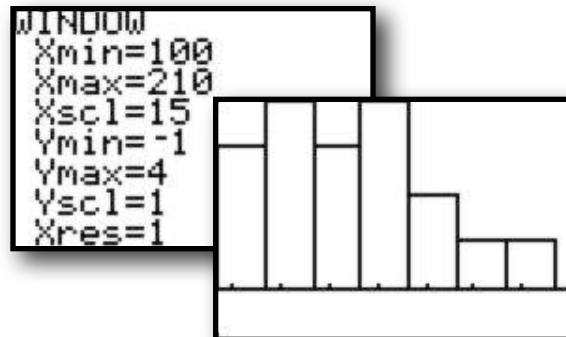
Comparing Data Displays

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]
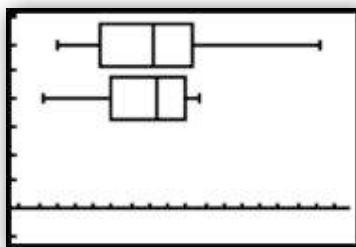   Your lists have now been stored for future reference.

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!

AP® Examination Tips

When taking the Advanced Placement Statistics Exam, you will most likely be asked to perform an exploratory data analysis. Remember, the calculator can be used to automate your calculations and provide basic data displays...however, it is your job to provide the contextual interpretation! Never answer a question by just copying a calculator plot or by simply listing the 1-variable statistics...be sure to label and interpret your analysis!

When making a plot:
- Be careful inputting your data
- Choose an appropriate plot
- Use the modified boxplot if you want to see if outliers exist
- Sketch the plot and LABEL axes!
- Interpret the SOCS of the graph in the context of the problem
- For comparisons, be sure to label each dataset on your plot

When calculating numeric summaries:
- Be careful inputting your data
- Choose the appropriate measures of center and spread
  - Mean and standard deviation, or 5-number summary
  - Be sure to refer to the sample standard deviation s_x, not
- Interpret the measures in the context of the problem
- When comparing datasets, be sure to compare the center and spread for each dataset in the context of the problem
- Know how to use the 1.5 IQR rule to determine outliers
- Be able to justify outliers on a modified boxplot by using this rule

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**
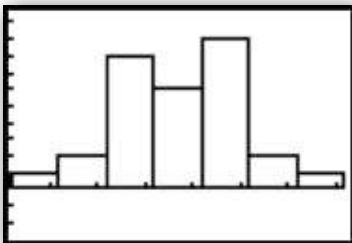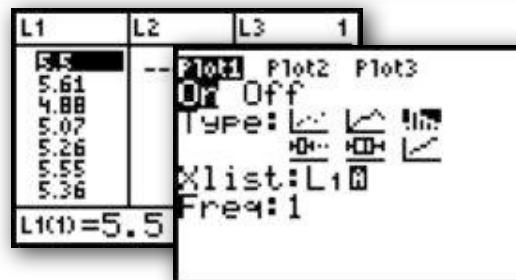
Note the shape of this distribution. It appears it can be approximated by a **normal distribution** with mean = 5.45 and standard deviation = 0.22.
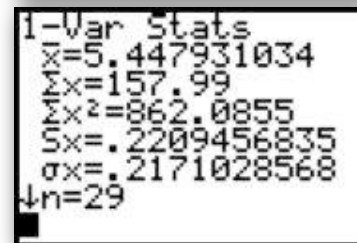
Remember, just because data appears normally distributed, that doesn't make it so. We can assess the normality of data by constructing a different stat plot on our calculator. This plot, known as a **normal quantile plot** or **normal probability plot** can be used to assess the adequacy of a normal 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.
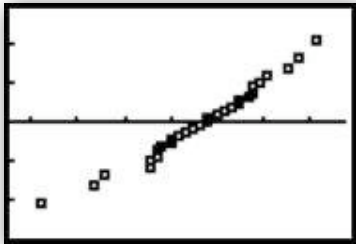
Assessing Normality

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.
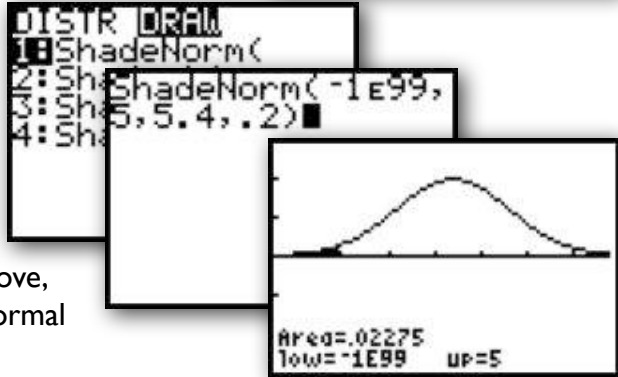
What proportion of density observations fall below 5?
  1. Press **2ND** [VARS] **(DISTR)**
  2. Choose **2:normalcdf(**
  3. Enter **normalcdf(-1E99, 5, 5.4, .2)**

```
DISTR DRAW
1:normalpdf(
2:normalcdf(
3:invNorm(
4:invT(
5:tpdf(
6:tcdf(
7↓X²pdf(
```

```
normalcdf(-1E99,
5,5.4,.2)
          .022750062
```

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)**

```
DISTR DRAW
1:ShadeNorm(
2:Sh
3:Sh
4:Sh
```
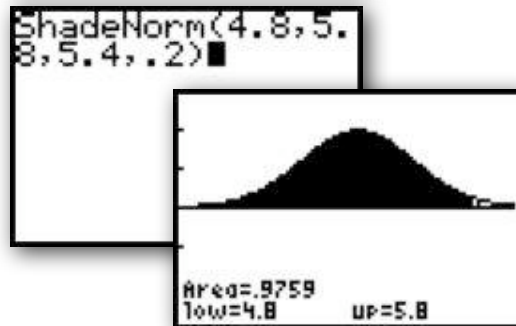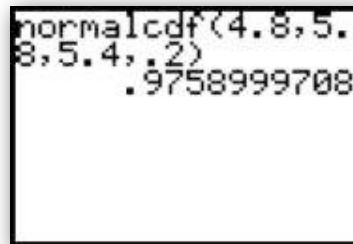
```
ShadeNorm(-1E99,
5,5.4,.2)█
```

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.

```
Area=.02275
low=-1E99    up=5
```

Find the proportion of density observations above 5.6:

```
normalcdf(5.6,1E
99,5.4,.2)
          .1586552596
```

```
ShadeNorm(5.6,1E
99,5.4,.2)
```

```
Area=.158655
low=5.6 99    up=1E99
```
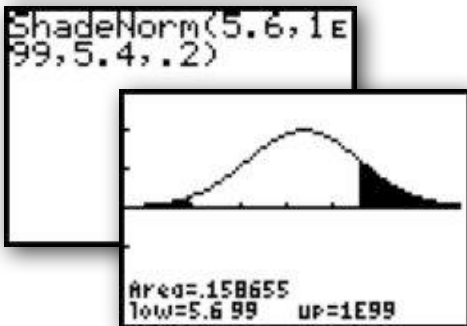
Find the proportion of density observations between 4.8 and 5.8:

```
normalcdf(4.8,5.
8,5.4,.2)
          .9758999708
```

```
ShadeNorm(4.8,5.
8,5.4,.2)█
```

```
Area=.9759
low=4.8    up=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)**

```
DISTR DRAW
1:normalpdf(
2:normalcdf(
3:invNorm(
4:invT(
5:tpdf(
6:tcdf(
7↓X²pdf(
```

```
invNorm(.8,5.4,.
2)
        5.568324247
```

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.

```
invNorm(.9)
        1.281551567
```

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

AP® Examination Tips

When taking the Advanced Placement Statistics Exam, you will most likely be asked to perform standard normal calculations. Remember, the calculator can be used to automate your calculations and provide basic data displays…however, it is your job to provide the contextual interpretation! Never answer a question by just copying a calculator plot or by using "calculator" speak.

When assessing normality
- Be careful inputting your data
- Choose the appropriate plot(s) (histogram and normal probability plot)
- Sketch the plot(s) and LABEL the axes!
- Interpret the graphs in the context of the problem
- Remember, linearity in the normal probability plot suggests normality

When performing normal calculations
- Sketch your normal distribution
- Label the mean and indicate the standard deviation
- Plot your point(s) of interest
- Shade the area you wish to calculate
- Make sure you choose **normalcdf(** _NOT_ **normalpdf(**
- Interpret the area in the context of the problem
- DO NOT USE calculator speak (eg: Don't write "*normalcdf = .023*" …

  Instead, say something like "*2.3% of observations fall below…*")
- Remember, sometimes it may be easier/faster to calculate a z-score by hand and use the table to find the appropriate area.

Chapter 3: Examining Relationships

3.1 Scatterplots and Correlation

In chapter 3 of YMS, we extend our exploratory data analysis procedures from univariate —or one variable—statistics to include bivariate—or two variable—analyses. Since many statistical studies involve analyzing the relationship between two variables, we must establish ways to display and quantify those relationships. We learn in this chapter that a **scatterplot** is an effective method of display while **correlation** and the **least-squares regression line** are effective quantitative descriptions of bivariate data. *Note: Like the previous chapters, be sure you understand the concepts behind these methods before using the calculator!*

Constructing a Scatterplot

To construct a scatterplot on the TI, you must first input the ordered pairs into your list editor. Typically, you will want to enter your explanatory variable into **L1** and your response variable into **L2**. However, you can enter the data into any available list—just be sure to remember where you have stored it!

Consider the following data on the number of manatees (large, gentle sea creatures) that are killed by boats along with the number of powerboat registrations (in thousands) in Florida from 1977-1990:

Registration	447	460	481	498	513	512	526	559	585	614	645	675	711	719
Manatees	13	21	24	16	24	20	15	34	33	33	39	43	50	47

1. Enter the data into **L1** and **L2**. Be careful to maintain ordered pairs!
2. Store these lists for later use:
   L1 [STO▸] **PBREG**   **L2** [STO▸] **MANTE**

3. Set up your **STAT PLOT** for a scatterplot of **L1** and **L2**.
4. Use **ZOOM STAT** to see the scatterplot.
5. Interpret the relationship in the context of the problem. {eg. *"There appears to be a strong, positive, linear relationship between powerboat registrations and the number of manatees killed."*}
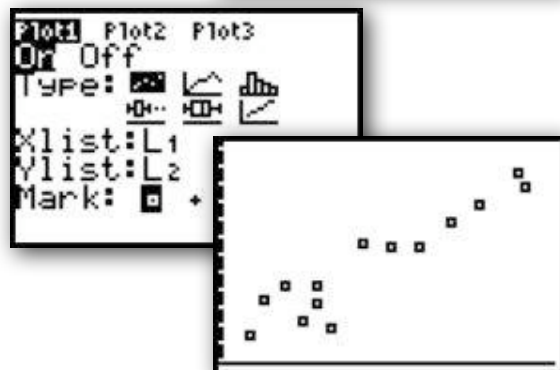
3.2 Correlation and Least Squares Regression
While the scatterplot provides a visual of the overall pattern between two variables, a numeric analysis is necessary for a better understanding of the relationship. To quantify the relationship, we use correlation and the least-squares regression line. Both of these numeric summaries involve rather tedious calculations. Thankfully, our calculators automate those calculations for us. As usual, be sure you understand the mechanics behind their derivation before using the calculator!

Setting Up Your Calculator to Display the Correlation Coefficient
Your TI will calculate and display both the equation for the least-squares regression line and the correlation coefficient with one command. However, you may need to adjust one setting to assure the calculator will display the correlation coefficient.

1. Press **2ND** ⓪ (**CATALOG**)
2. Select "**DiagnosticOn**"
3. Press ⌷ENTER⌷

Your calculator is now set to display the correlation coefficient.

Calculating Correlation and the Least-Squares Regression Line
Suppose we want to calculate the line of best fit for our powerboat, manatee data. We have already seen that there appears to be a strong, positive, linear relationship between these variables. Finding the least-squares regression line will provide us with a model to predict manatee deaths based on powerboat registrations.

1. Enter the data into **L1** and **L2**.
2. Press **STAT CALC 8:LinReg(a+bx)**
3. On your homescreen, indicate **LinReg(a+bx) L1, L2**
4. Press ⌷ENTER⌷
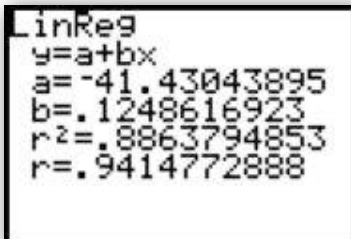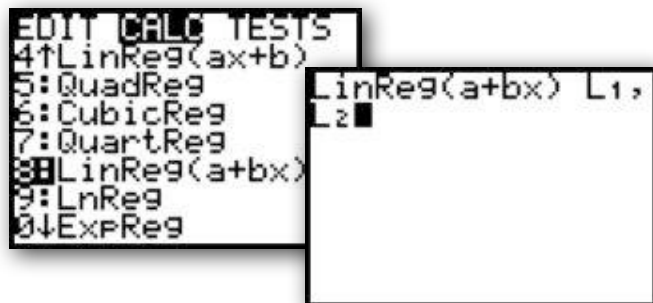
Interpret the output in the context of the problem. The printout indicates the equation of the least squares regression line is: **manatees=-41.43+0.12(powerboats)**. Further, the correlation coefficient **r=0.94** supports our initial conclusion that there is a strong positive, linear relationship between these two variables.

Graphing the Least-Squares Regression Line on the Scatterplot

If you would like to see how well the LSRL fits the data, you can graph it simultaneously with the scatterplot. When you calculate a LSRL, it is stored in the calculators "$\boxed{\text{VARS}}$" menu. To graph it, you simply need to paste it into **Y1**.

1. **ZOOM STAT** to a scatterplot of the data.
2. Calculate the LSRL
3. Press **Y=** $\boxed{\text{VARS}}$ **5:Statistics…**
4. Choose **EQ 1:RegEQ**
5. Press **GRAPH**
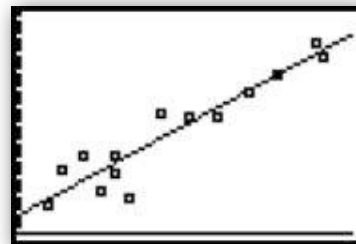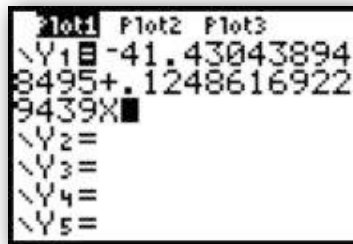
Creating a Residual Plot

Notice the observed values don't fall perfectly on the LSRL. Each prediction by the LSRL has a "residual" value: residual=observed-predicted. Plotting the residuals gives us a better idea about the appropriateness of a linear model for our data. If a linear model is reasonable, a residual plot should show no discernable pattern.

1. Calculate the LSRL for your data
2. Set **STAT PLOT** for a scatterplot
3. Set **Xlist**: to **L1**
4. Set **Ylist**: to **2ND STAT (LIST) "RESID"**
5. **ZOOM STAT** to see the residual plot
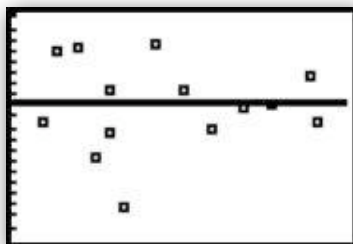
Chose **TRACE** to see individual residual values for each x. Note, there is no discernable pattern in this plot, suggesting the least-squares regression line may be an appropriate prediction model.

15

AP® Examination Tips

When taking the Advanced Placement Statistics Exam, you will most likely be asked to find and interpret a least-squares regression line. Remember, the calculator can be used to automate your calculations and provide basic data displays…however, it is your job to provide the contextual interpretation! Never answer a question by just copying a calculator plot or by using "calculator" speak.

When finding a least squares regression line
- Be careful inputting data. Be sure to maintain ordered pairs.
- Remember which list is "explanatory – x" and "response – y."
- Begin by assessing the strength, direction, and form of a scatterplot.
- Don't forget to tell the calculator which lists you'd like the LSRL for.
- Remember to write the equation in context.

When finding a residual plot
- Be sure to find the least squares regression line first.
- Remember, the residuals are stored in **RESID** in the **List** menu.
- Remember, no pattern suggest the linear model may be appropriate.

When finding/interpreting r and r²
- If you don't see r and r^2, turn "**Diagnostics On**"
- Interpret r by stating the strength and direction of the linear relationship in context.
- r^2 is interpreted as "*The percent of variability in predicted y-values that can be explained by the least squares regression model on x.*"
- Remember, strong r and r^2 values alone do not justify use of the linear model, be sure to look at your residual plot!

Chapter 4: More About Bivariate Relationships

4.1 Transforming to Achieve Linearity

In chapter 4, we learn how to find nonlinear models for data by transforming to achieve linearity. The transformation process can be very time consuming, however, we can use our calculator to automate some of our work. It is unlikely you'd be asked to find a power or exponential model on the AP exam. However, knowing the procedure will lead to a better understanding of regression models. *Note: Like the previous chapters, be sure you understand the concepts behind these methods before using the calculator!*

Modeling Exponential Growth

Consider the following data on the number of transistors on a chip. According to "Moore's Law," the number grows exponentially over time. {In 1965, Gordon Moore, a co-founder of Intel, actually predicted the number would double every 18 months.}

| Year	1971	1972	1974	1978	1982	1985	1989	1993	1997	1999	2000
Transistors	2250	2500	5000	29000	120000	275000	1180000	3100000	7500000	24000000	42000000

1. Enter the data into **L1** and **L2**.
2. **ZOOM STAT** to a Scatterplot
3. Calculate the LSRL
4. View the Residual Plot
5. Press ENTER

Scatterplot (L1,L2)

Note the relatively low coefficient of correlation and the distinct pattern in both the scatterplot and residual plot. We have evidence to suggest a nonlinear model may be more appropriate.

To find the "best fitting" nonlinear model, we must transform our data to achieve linearity. To do this, define **L3** to be ln(x) and **L4** to be ln(y). As we learned in Chapter 4, if (x, ln(y)) is linear, then an exponential model may be appropriate. If (ln(x),ln(y)) is linear, a power model may be appropriate.
Use the list editor to transform your data. Then, **ZOOM STAT** to plots of (**L1, L4**) and (**L3,L4**).
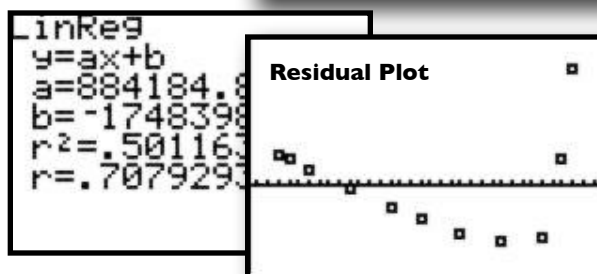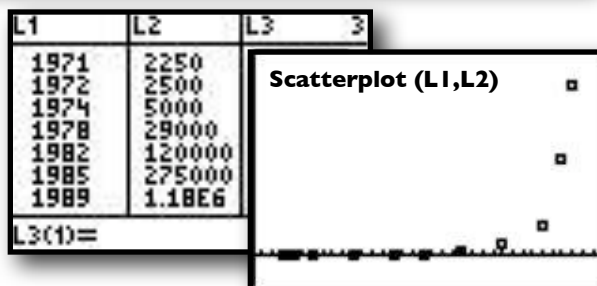
Residual Plot

(L1,L4) = (x, ln(y))

(L3,L4) = (ln(x), ln(y))
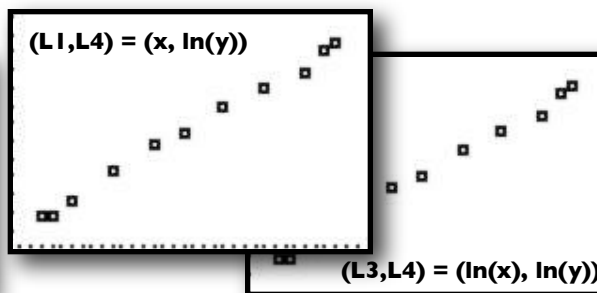
Notice, both (x, ln(y)) and (ln(x), ln(y)) are approximately linear. It appears either an exponential or a power model could be used. Since the situation appears to be exponential in nature, we will proceed to build an exponential model.

1. Find the LSRL of (x, ln(y)).
2. Transform (ln(y))-hat = a+bx back to the original (x,y) context.
3. y-hat = $e^{(a+bx)}$
4. *transistor-hat* = $e^{(-645.87+.331(year))}$

Note: **STAT CALC ExpReg L1,L2** can also be used to find this model, but may result in an "overflow" error.

Modeling Power Growth
The process for modeling power growth is the same as that for modeling exponential growth.

1. Enter your data and check the scatterplot.
2. If linear, find the LSRL. If not linear...
3. Transform your data by defining **L3=ln(L1) and L4=ln(L2)**.
4. Check both **(L1, L4)** and **(L3, L4)** scatterplots for linearity. Consider *r* and residual plots if necessary. Linear **(L1, L4)** implies exponential. Linear **(L3,L4)** implies power model.
5. Find the LSRL of the 'more linear' transformed data.
6. Transform the Linear Regression Equation back to the original data.

If **(L1, L4)** was more linear, ln(y)-hat = a + bx

Exponential Model: y-hat = $e^{(a+bx)}$

If **(L3, L4)** was more linear, ln(y)-hat = a + b(ln(x))

Power Model: y-hat = $e^a x^b$

While the TI is capable of finding an exponential or power model through STAT CALC, it helps to know *how* it goes about doing so. Again, the chances are minimal that you'd have to go through this entire procedure on the AP Exam. Make sure you understand the process as you may be asked to describe it or justify the use of a particular model.

AP® Examination Tips

When taking the Advanced Placement Statistics Exam, the chances are slim that you would need to perform a transformation to achieve linearity. You may be asked to interpret the results of such a transformation, or to justify the use of a particular regression model based on graphical and numeric output. When doing so, be sure to interpret the output in the context of the problem!

When finding a non-linear model
- Start with an analysis of the original (x,y) data
- Check BOTH (x, ln(y)) and (ln(x), ln(y))
- Refer to *r* and residual plots when determining which is "more linear"
- Be sure to keep your lists straight…don't forget which is ln(x) or ln(y)
- LABEL sketches with appropriate variables!

When interpreting results
- Keep your variables straight…are you referring to y or ln(y)?
- Remember, linear (x, ln(y)) implies exponential while linear (ln(x), ln(y)) implies a power model may be more appropriate.
- Be careful when using models to "plug in" appropriate measures and to interpret predicted values correctly.
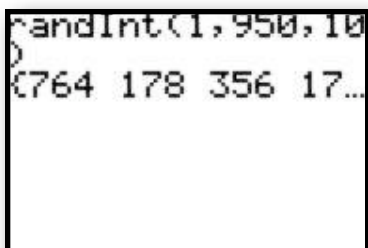- Be careful not to extrapolate too far beyond the observed data.

Chapter 5: Producing Data

5.1-5.2 Designing Samples and Experiments

In Chapter 5 of YMS, we are introduced to the second major theme in AP Statistics--Collecting/ Producing Data. There is not a lot we need to do on our calculators in this chapter. However, they can come in handy when selecting simple random samples. Rather than relying on a table of random digits, we can use the pseudo-random generator in the calculator to select a sample.

1. Number the individuals in the population you wish to sample from 1 to *n*.
2. Push MATH **PRB 5:randInt(**
3. Execute **randInt(1,n, sample size)**

For example, suppose I wanted to select an SRS of 10 individuals from a population of size 950. I would number each individual from 1 up to 950, then execute **randInt(1,950,10)**
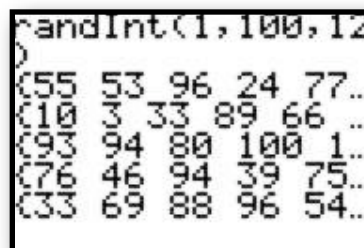
Scrolling to the right will show the rest of the randomly selected numbers. Note: repeats are possible, so you may want to select more than 10 and then take the first 10 non-repeated numbers.

This process can also be used to randomize in experimental design. Executing **randInt(1,2)** can be used to simulate flipping a coin to split individuals in to two experimental groups or repeating the process above can be used to randomly select a treatment group.

5.3 Simulating Experiments

In some cases, actually carrying out an experiment can be slow, costly, or logistically difficult. We can use a probability model and our TI to calculate a theoretical answer for some situations. The key to using the calculator is to start with a well-defined model that reflects the the situation we are studying. For example, suppose 90% of passengers show up for a given flight. To ensure flights are as full as possible, airlines often overbook, hoping some passengers don't show up and all who do get a seat. What is the probability a 10 seat plane will be overbooked if 12 tickets are sold?

1. Define a probability model.
   - Select a number between 1 and 100.
   - 1-90 = show up for the flight
   - 91-100 = "no show"
2. Draw 12 random numbers between 1 and 100.
3. Count the "no shows"
4. Keep track of overbooked vs. ok
5. Repeat the process and estimate the probability.

AP® Examination Tips

When taking the Advanced Placement Statistics Exam, you will most likely be asked to design an experiment or answer questions related to a sampling process or

When describing randomization:
- Be complete when describing the process used to select a sample or determine experimental groups…DO NOT simply write the calculator command you used. For example, write "*I numbered the individuals from 1-20 and used a random generator to select 5 for my sample.*" NOT "*I typed randInt(1,20,5)*"
- Make sure you understand WHY we randomize in survey situations or experimental design.

When simulating an experiment:
- Define your probability model. That is, indicate what numbers stand for what situations. For example, write "*To simulate choosing individuals from a pool in which 70% are employed, I will randomly select numbers between 1 and 100. 1-70 represent employed individuals while 71-100 represent unemployed individuals.*"
- Repeat the simulation a number of times and keep track of your results so you can calculate the theoretical probability.

Chapter 6: Probability-The Study of Randomness

This chapter focuses on the theory and formulas governing probability and the study of randomness. Other than using your calculator to simulate experimental situations, there is little else in this chapter that is dependent upon the TI. You may wish to use your calculator to illustrate the Law of Large Numbers by simulating repeated flips of a fair coin. Your TI can simulate flipping the coin repeatedly and then graph the cumulative frequency of heads or tails.

1. Press **2nd STAT OPS 5:seq**
2. Execute **seq(X,X,1,150)** STO▶**L1** to generate a sequence of integers from 1 to 150 in **L1**.
3. Execute **randInt(0,1,150)** STO▶ **L2** to simulate flipping a coin 150 times.

4. Press **2nd STAT OPS 6:cumSum(**
5. Execute **cumSum(L2)** STO▶ **L3** This stores the cumulative number of heads (or tails) in list 3.
6. Execute **L3/L1** STO▶ **L4** to calculate the relative frequency of heads (or tails) up to each number of flips.
7. Set your **StatPlot** to create an xyLine plot of **L1, L4**

8. **ZOOM STAT** to see how the relative frequency of heads (or tails) settles down as the number of trials increases.
9. Plot **y=.5** to see how the relative frequency of heads (or tails) approaches .5 as the number of trials increases.

Chapter 7: Random Variables

In this chapter of YMS, we are introduced to the concept of a random variable as a numerical variable whose values are determined by the outcome of a random phenomenon. Further, we are introduced to rather tedious formulas for calculating the mean and variance of random variables.