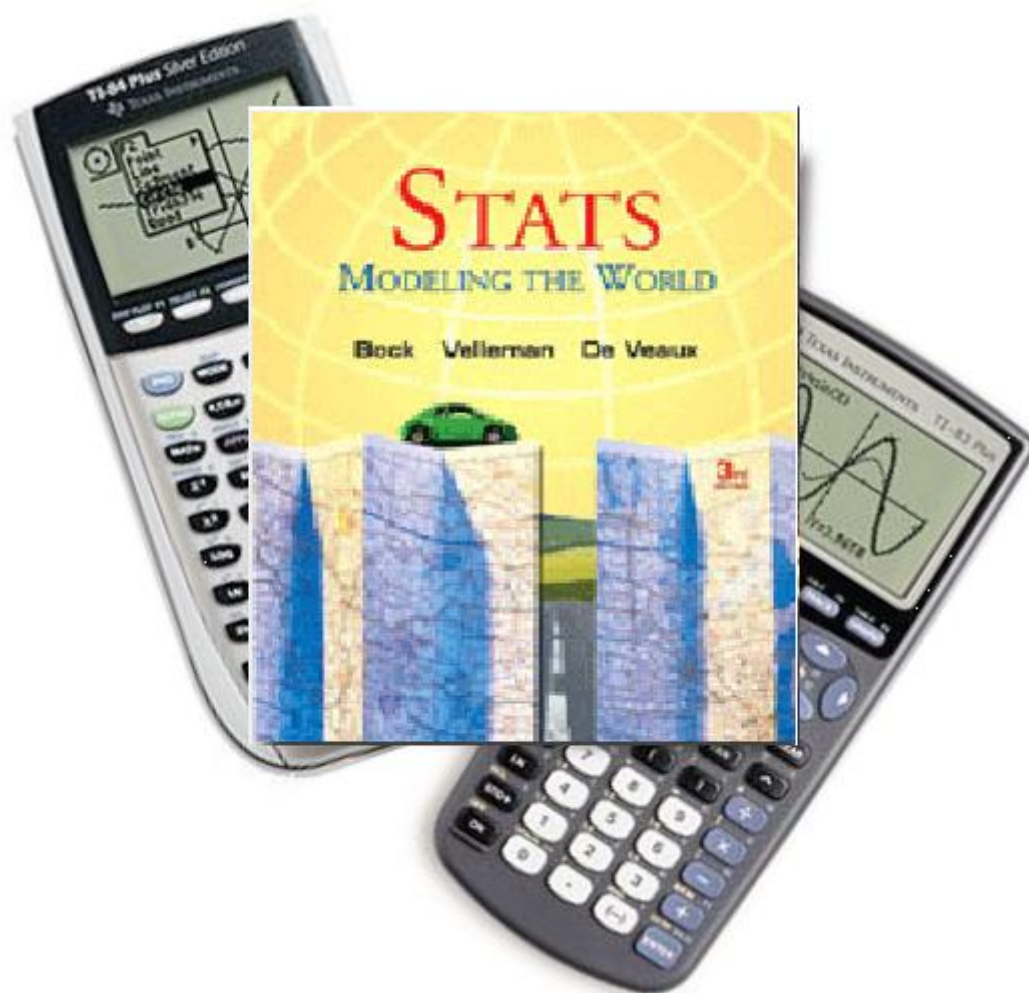


AP Statistics: TI83+/84+ User Guide

**Oegie Briones
Concord High**



AP Statistics

AP STATISTICS: CALCULATOR NOTES

These calculator notes should all be compatible with the functions of the TI-83, TI-83+, TI-84 and TI-84+ graphing calculators. Any reference numbers that refer to chapters and sections are not accurate, as they refer to the old textbook (*The Practice of Statistics*, 3rd edition) we used several years ago.

DESCRIBING DATA WITH GRAPHS

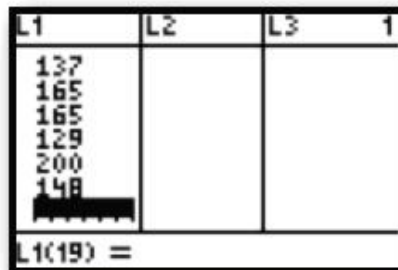
Displaying Univariate Data

Consider the following data on the Survey of Study Habits and Attitudes (SSHA) score for 18 female college students. The test evaluates motivation, study habits, and attitude 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 **2nd**, **Y=** (**STAT PLOT**)
2. Select 1: Plot1...
3. Press **ENTER** to highlight **On**
4. Select the **Histogram** option under **Type**:

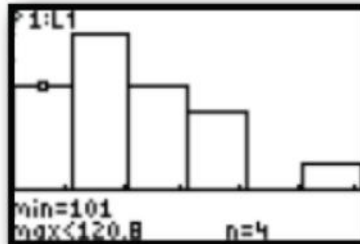


AP Statistics

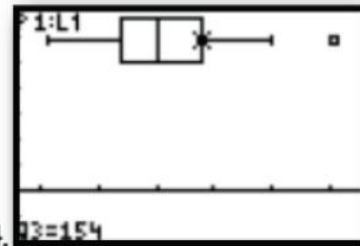
You are now ready to view a histogram of the "SSHA" data. To do this, you need to set your window to the appropriate values. You can do this by changing the parameters in the **WINDOW** mode, or you can **Zoom** directly to the data.

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

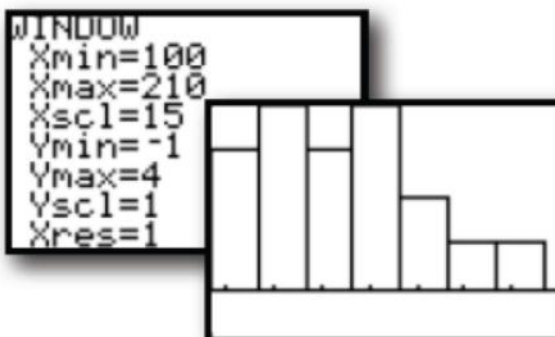


Using the "SSHA" data, select **modified boxplot** to get another view of the distribution.



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



AP Statistics

DESCRIBING DATA WITH NUMBERS

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

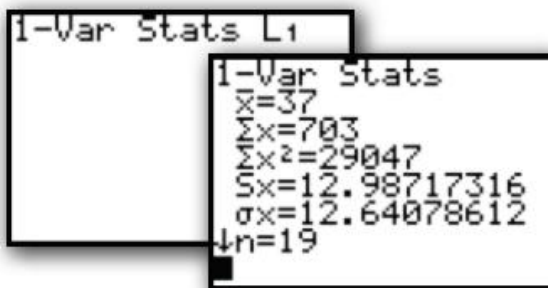
The calculation of each of these measures, especially the standard deviation, can be quite tedious. Thankfully, the TI can automate those calculations for us. Like plotting data, the calculator requires that you enter the dataset before it can report a numeric summary.

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 Statistics

THE NORMAL DISTRIBUTION

The Normal Probability Plot, which is covered below, is useful but not necessary for our class. If you're handy with your calculator, this can be a useful tool when we get to inferential statistics.

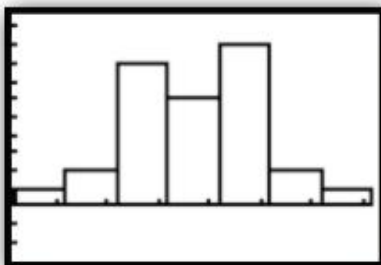
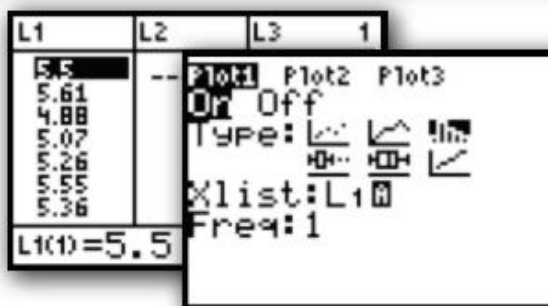
Approximating a Normal Curve

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.

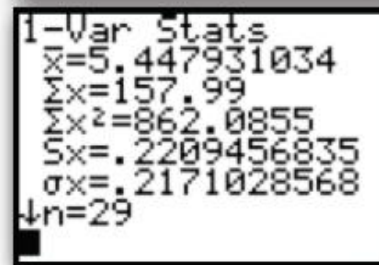
Density of the earth (as a multiple of the density of water)

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.

AP Statistics

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.

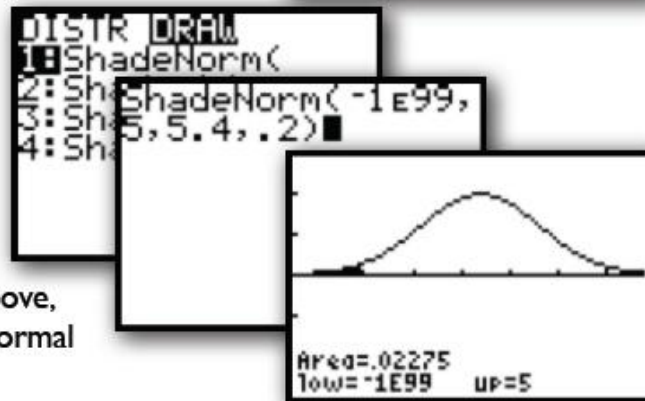
What proportion of density observations fall below 5?

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:

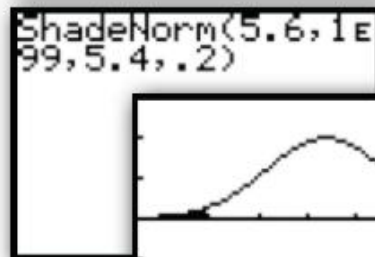
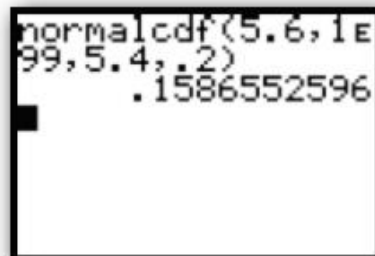
1. **2ND** **[VARS]** (**DISTR**)
2. Choose **DRAW 1:ShadeNorm**(
3. Enter **ShadeNorm(-1E99, 5, 5.4, .2)**



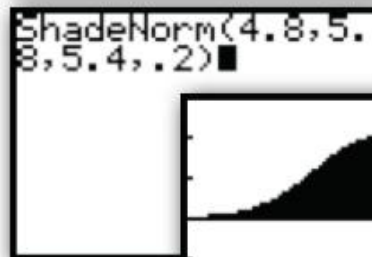
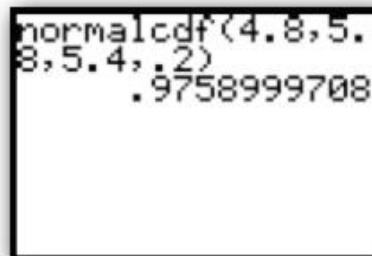
Notice in both cases we find approximately 2.3% of observations will fall below a measurement of 5.

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:



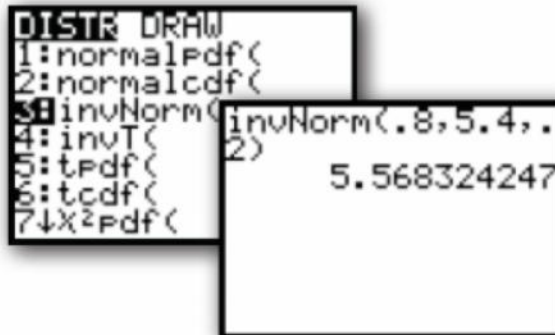
AP Statistics**Inverse Normal Calculations**

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**

AP Statistics

SCATTERPLOTS, CORRELATION, AND REGRESSION

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 $\text{STO} \rightarrow$ **PBREG** **L2** $\text{STO} \rightarrow$ **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."}



AP Statistics**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**

```
EDIT CALC TESTS
4↑LinReg(ax+b)
5:QuadReg
6:CubicReg
7:QuartReg
8↓LinReg(a+bx)
9:LnReg
0↓ExpReg
```

```
LinReg(a+bx) L1,
L2
```

```
LinReg
y=a+bx
a=-41.43043895
b=.1248616923
r2=.8863794853
r=.9414772888
```

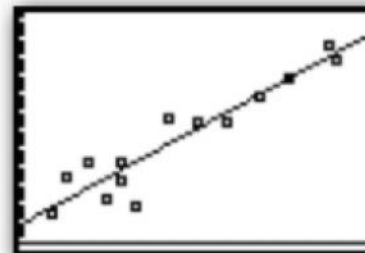
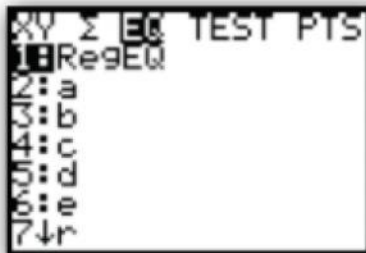
Interpret the output in the context of the problem. The printout indicates the equation of the least squares regression line is: **manatees=-41.43+0.12(powerboats)**. Further, the correlation coefficient **r=0.94** supports our initial conclusion that there is a strong positive, linear relationship between these two variables.

AP Statistics

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 "VARS" menu. To graph it, you simply need to paste it into **Y1**.

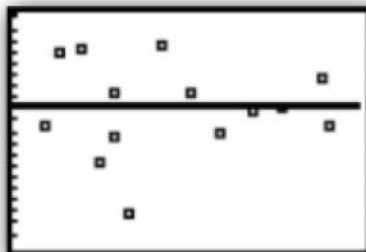
1. **ZOOM STAT** to a scatterplot of the data.
2. Calculate the LSRL
3. Press **Y=** **[VARS]** **5:Statistics...**
4. Choose **EQ 1:RegEQ**
5. Press **GRAPH**



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.

AP Statistics

**SCATTERPLOTS CONTINUED:
TRANSFORMING NONLINEAR SCATTERPLOTS**

4.1 Transforming to Achieve Linearity

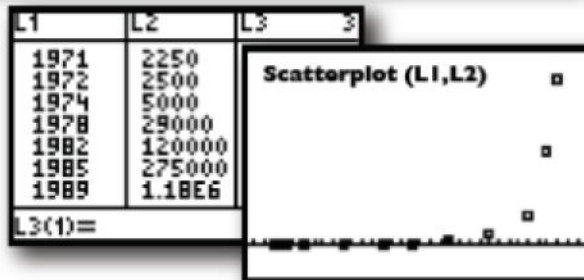
In chapter 4, we learn how to find nonlinear models for data by transforming to achieve linearity. The transformation process can be very time consuming, however, we can use our calculator to automate some of our work. It is unlikely you'd be asked to find a power or exponential model on the AP exam. 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

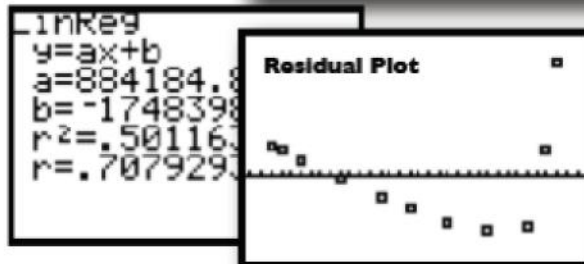
Consider the following data on the number of transistors on a chip. According to "Moore's Law," the number grows exponentially over time. {In 1965, Gordon Moore, a co-founder of Intel, actually predicted the number would double every 18 months.}

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**

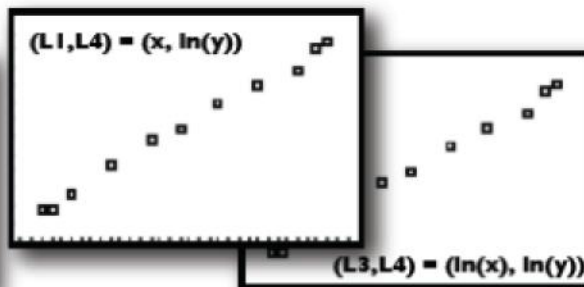


Note the relatively low coefficient of correlation and the distinct pattern in both the scatterplot and residual plot. We have evidence to suggest a nonlinear model may be more appropriate.



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).

L2	L3	L4
2250	7.7187	7.7187
2500	7.924	7.924
5000	8.5172	8.5172
29000	10.275	10.275
120000	11.695	11.695
275000	12.525	12.525
1.18E6	13.981	13.981



AP Statistics

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 $(\ln(y))\text{-hat} = a+bx$ back to the original (x,y) context.
3. $y\text{-hat} = e^{(a+bx)}$
4. $\text{transistor-hat} = e^{(-645.87+331(\text{year}))}$



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, $\ln(y)\text{-hat} = a + bx$

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

If **(L3, L4)** was more linear, $\ln(y)\text{-hat} = a + b(\ln(x))$

Power Model: $y\text{-hat} = e^a x^b$

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.

RANDOM NUMBERS AND SIMULATIONS

- To generate random digits using your calculator, and conduct a random sample.

- Number the individuals in the population you wish to sample from 1 to n .
- Push **MATH** **PRB** 5:randInt(
- 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?

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