

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!*

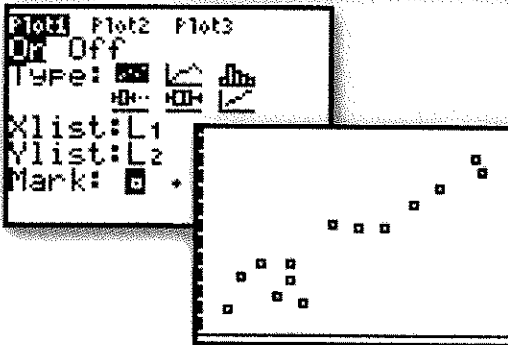
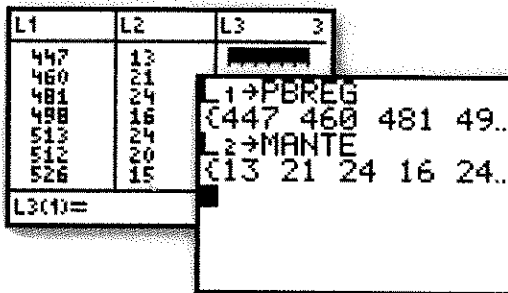
Constructing a Scatterplot

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



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**

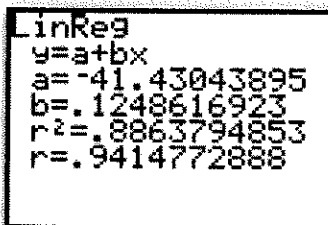
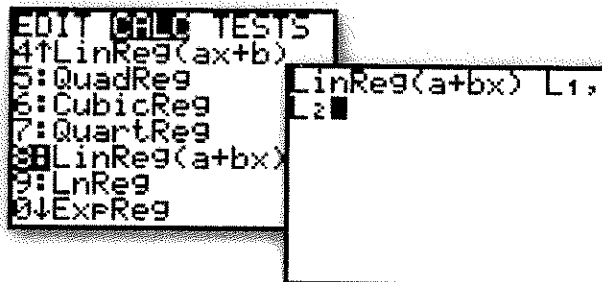


Your calculator is now set to display the correlation coefficient.

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**

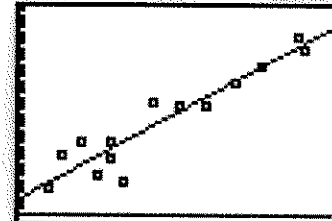
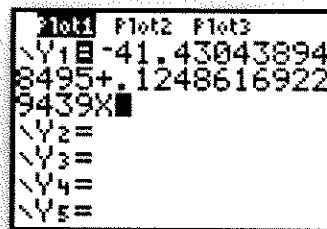
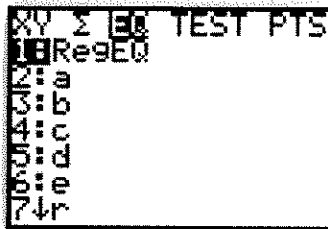
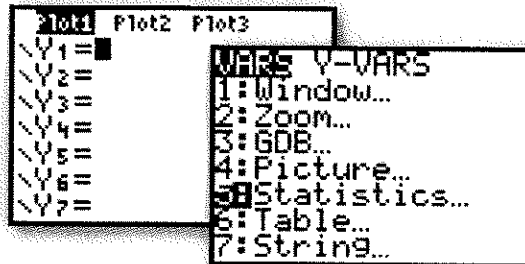


Interpret the output in the context of the problem. The printout indicates the equation of the least squares regression line is: **manatees=-41.43+0.12(powerboats)**. Further, the correlation coefficient **r=0.94** supports our initial conclusion that there is a strong positive, linear relationship between 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 “**VARS**” menu. To graph it, you simply need to paste it into **Y1**.

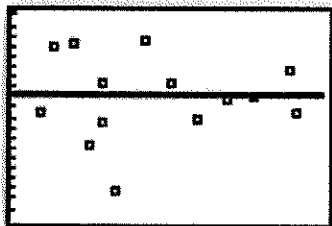
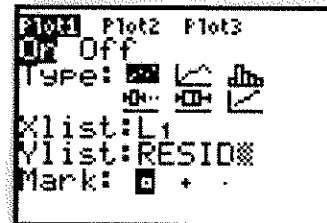
1. **ZOOM STAT** to a scatterplot of the data.
2. Calculate the LSRL
3. Press **Y=** **VARS** **5:Statistics...**
4. Choose **EQ 1:RegEQ**
5. Press **GRAPH**



Creating a Residual Plot

Notice the observed values don't fall perfectly on the LSRL. Each prediction by the LSRL has a “residual” value: $\text{residual} = \text{observed} - \text{predicted}$. 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[®] 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 r^2

- 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!

