What is the purpose of linear regression? (linear regression – regress back to the data and use it to form a line) Some data exhibit a relationship between two variables; graphing the ordered pairs, at times, indicates a nearly linear relationship. If this is the case, a “line” can represent the relationship for predictive purposes. These statements will become understandable as you examine the following example.

The following data are two sets where the individual pieces are related to other individual pieces in the form of ordered pairs. In real life, for the purpose of best estimating what the weight of a normal-developing individual baby should be, a SRS of hundreds of infants would be used for the model. With such a model, a doctor might be able to determine if a child is progressing normally. For reasons of simplification, the following data will be used to develop the procedures you must use on your calculator to get the “line-of-best-fit” or “regression line”.

<table>
<thead>
<tr>
<th>Age (days)</th>
<th>5</th>
<th>10</th>
<th>15</th>
<th>20</th>
<th>25</th>
<th>30</th>
<th>35</th>
<th>45</th>
<th>55</th>
<th>65</th>
<th>75</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weight (mean; oz.)</td>
<td>111</td>
<td>121</td>
<td>130</td>
<td>138</td>
<td>150</td>
<td>162</td>
<td>170</td>
<td>195</td>
<td>119</td>
<td>230</td>
<td>252</td>
</tr>
</tbody>
</table>

Which variable is the “independent’ variable, the one that will change no matter what the other variable does? ________________________ This is also called the ______________ or ____________________ variable. The other variable is the “dependent” variable; it is also called the ____________________ variable.

Plot this data on a set of axes (on graph paper) using the explanatory / predictor variable as the x-axis. Use a scale of “5” for x (0, 5, 10, 15, … 75, 80, 85) and a scale of “15” for y (0, 15, 30, 45, … 240, 255, 270).

What is the “trend” of the data? ________________________ The paired data seems to be what type of “geometric” relationship? It is almost a ___________________. Lightly draw a line that “fits” the data.

Estimate the slope and y-intercept of the line: m ≈ _________ y-intercept ≈ __________ (Be careful finding the slope; your axes scales are different!)

Put the age in one list, L₁, and the weight in another list, L₂, and make a __________________________ plot.

(Press Y= and delete or turn off any graphs. Go to 2nd Y= and ENTER on the first plot; press ENTER again to turn it on. Use the arrow keys to go to the first plot and press ENTER; it’s called a “scatter plot”. Put in the proper lists, L₁ and L₂, and choose a “mark”. Press ZOOM 9, which is ZOOM STATISTICS. You should see a scatter plot of the points just like you had on graph paper. Press TRACE; using your arrow keys, you can see the ordered pair for each point.)

Let’s use the calculator to get its line-of-best-fit so that everyone has the same line. What is the calculator’s equation or line-of-best-fit? Write it in the context of the data or problem.

(Go to the home screen by pressing 2nd, MODE (QUIT), press STAT, CALC; choose #4 or #8, depending on what your instructor wants you to use. Put in your lists separated by a comma (above the 7) and press ENTER; yours might look like this: LinReg(ax + b) L₁, L₂) What is the calculator’s slope? _________ What does it mean in the context of the problem?

Age, predictor or explanatory variable; weight = response (criterion) variable; upward / increasing; line; m is about 2; y-intercept is about 100; scatter plot; y = 2.054 x + 99.78; in the context of the problem, we are trying to predict the weight from the age in days, so the answer is Wt. (oz.) = 2.054 Age (days) + 99.78; m = 2.054; slope is a rate; the average weight is changing at about 2 ounces per day.
Focus – Linear Regression (M202)

How can the line be transferred to your physical scatter plot? First you need to graph the correct equation on the calculator. You could do this using the Y= menu; this would require that you round some of the numbers unless you wanted to put all the decimal places in the equation. There is an easier way. Redo the linear regression in this way: After your LinReg(ax + b) \( L_1, L_2 \) put in a comma and press VARS, scroll over to Y-Vars; press ENTER on Function and ENTER on \( Y_1 \); your home screen should look like this: LinReg(ax + b) \( L_1, L_2, Y_1 \). Press enter and you get the same thing you had before; now press GRAPH. Wow! The line is now on the scatter plot!

What is the y-intercept? In order to get this, looking at the graph, press TRACE. Now press the up or down arrow key to get the cursor on the line. Press “0” and you will get the y-intercept; it is \((0, \; \text{__________})\). What does it mean in the context of the problem?

If this child was 23 days old, what would you expect the child to weigh according to your projection or prediction line? ________ This process is called ________________________________.

(Go to the graph; press TRACE and up or down arrow key to get on the line. Press 23 ENTER and the weight is predicted to be 144.3 ounces. This process of predicting between known data points is called “interpolate or interpolation”.)

If this child was 2 months old, what would you expect to weigh according to your projection or prediction line? ________ … 6 months old? ________ This process is called ________________________________.

(2 months is about 60 days, so TRACE on the line at 60; did you get about 223.0 ounces? To get 6 months or 6x30 days, you will run off the graph, so change the WINDOW Xmax to 200 and press GRAPH; now trace on the line. Did you get about 510.5 ounces?) This process of predicting is called extrapolate or extrapolation.)

Is the last answer reasonable? 510 / 16 is about 32 pounds! What happened to our model?

___________________________

(510.5 ounces / 16 is about 32 pounds, that is unreasonable for a 6 month old! The model fell apart; it is only good for a limited set of data, or in this case, time in days. A new model will be needed for toddlers.)

When you did your correlation (co-relation) LinReg(ax + b) \( L_1, L_2 \), did you get an “r” and “r^2”? If you did not see values for “r” and “r^2”, press 2nd, “0” (CATALOG) and scroll down to “DiagnosticOn: and press ENTER’’ the screen should say “DONE”. Now set up your linear regression: LinReg(ax + b) \( L_1, L_2 \) and press ENTER and you will see the values of “r” and “r^2.”. What is the correlation or correlation coefficient? \( r = \; \text{__________} \) What is it telling you?

___________________________

What is \( r^2 \)? ________ What is it telling you? ________________________________

_________________________

What might be some “lurking variable”? ________________________________

What is 1- \( r^2 \)? ________What is it called? ________________________________

\( r = 0.998 \); the sign (+) tells you the data have an upward or increasing trend and the absolute value of \( r = 0.996 \) is close to 1, which tells you the data are very close to the line. \( r^2 = 0.997 \); this tells you your predictions between data points, interpolations, are very accurate! “Lurking variables” are factors that were not included in the original collection of the data; these might be the health of the infant, the amount of activity or exercise it gets, any illness it may have, and how well the baby is adjusting to the mother’s milk or formula. 1 - \( r^2 \) = 0.003 and is called the coefficient of determination (proportion of variance accounted for by the predictor variable).