One and Two Variable Statistics – Review
ID: 11489

Time Required
40-60 minutes

Activity Overview
In this activity, students will review the concepts that they have learned thus far in statistics. The first part of the activity includes one-variable topics such as graphing quantitative variables, calculating measures of central tendency and spread, and making comparisons. The second part includes two-variable topics including linear regression, correlation, and interpolation. The coefficient of determination will be introduced.

Topic: Review
- *One-variable statistics*
- *Two-variable statistics*

Teacher Preparation and Notes
- *This can be used as a review lesson or an assessment. Students must have a working knowledge of graphing histograms and box plots, spread, quartiles, scatter plots, lines of best fit, linear regression, and interpolation.*
- *Calculator skills – Students should be able to use the TI-84 to enter data in lists, graph and interpret histograms and box plots, graph scatter plots, and find lines of best fit using linear regression.*
- *Sample answers to the worksheet questions are at the end of the document.*
- *This activity uses data from the World Health Organization. Go to http://apps.who.int/ghodata/?vid=3200&theme=country for more data about each country and other countries.*
- ***To download the student worksheet and calculator file, go to education.ti.com/exchange and enter "11489" in the keyword search box.***

Associated Materials
- *1and2VariableReview_Student.doc*
- *1and2VariableReview_HealthData.xls*
- *HEALTH.8xp (program)*

Suggested Related Activities
To download any activity listed, go to education.ti.com/exchange and enter the number in the keyword search box.
- *Graph My Center (TI-84 Plus family) — 10895*
- *Box Plots and Histograms (TI-84 Plus family) — 8200*
- *Run Me a Linear Regression (TI-84 Plus family) — 5137*

.

Data

The program **HEALTH**, creates 6 lists named LIFE, CAL, INFAN, HEALT, WATER, and SANIT and sets these lists up in the List Editor.

The HealthData excel document contains this data. You may wish to display it using an overhead projector or interactive whiteboard for students to reference during the activity.
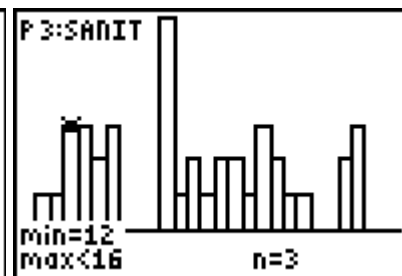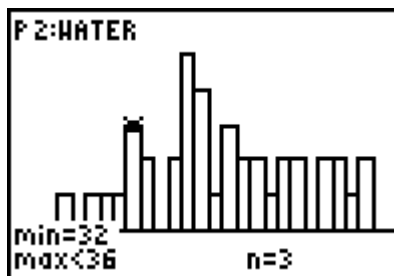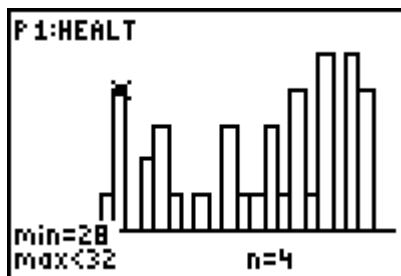
One-Variable Statistics

Students are asked to compare the three distributions that describe the percentages of people who have access to health services, safe water, and sanitation.

They will first make a histogram for each variable.

**Note:** They need to use the same window for each variable in order to make accurate comparisons.

Then, students are to make a series of three parallel box plots for these variables. All three plots need to be On. Press ZOOM and select **ZoomStat** to display the box plots.

Press 2nd [FORMAT] and select **AxesOff** to hide the axes.

Students can press TRACE to view the five-number summary. The left and right arrow will move through a box. The up and down arrow will move from box to box.
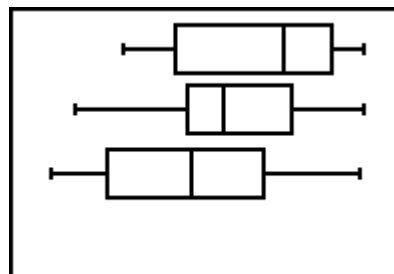
Two-Variable Statistics

Students are asked to investigate relationships between two variables, finding a positive correlation, negative correlation, and no correlation.

They will probably find the two variables that show these types of correlation through trial and error, but encourage them to think about how one variable might affect another. For example, one may think that if you drink better water, then you will have a healthier life and live longer, so life expectancy and safe water might have a correlation.

Students will answer Questions 8 through 13 using the scatter plot of infant mortality vs. life expectancy. They should have **INFAN** as Xlist, the independent variable, and **LIFE** as Ylist, the dependent variable.

After describing the relationship between the two variables, students will find their own line of best fit (STAT **> CALC > Manual-Fit**). Press ENTER to place the first point of the line, move the cursor to the other side of the plot, press ENTER to place the second point. The equation will appear in the top-left corner of the screen.

Students will find the linear regression equation. They can show the line on the scatter plot by selecting STAT **> CALC > LinReg(ax+b)** and entering ∟**INFAN,** ∟**LIFE, Y₁**. After calculating, press GRAPH.

This will also give students the values of r and r^2 if they have selected **DiagnosticOn** from the Catalog first.
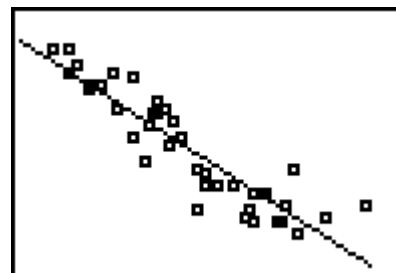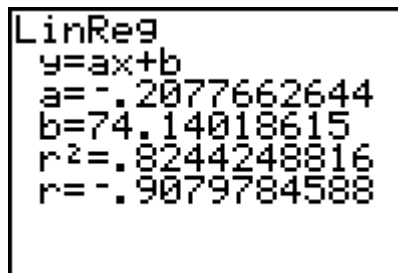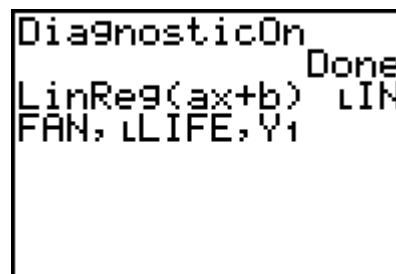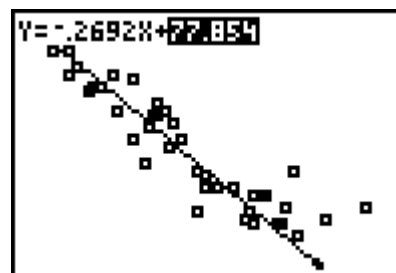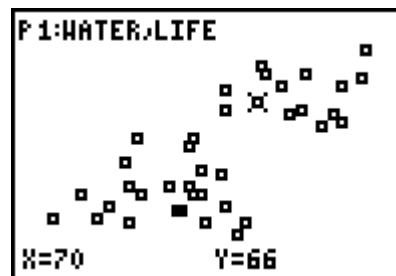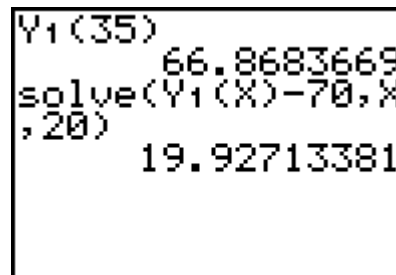
If using Mathprint OS:

To turn the DiagnosticOn, press MODE and arrow down to the second set of options. Press ENTER on **On** to the right of STAT DIAGNOSTICS: and then press 2nd [QUIT] to return to the Home Screen. This should be done before executing the Linear Regression command.
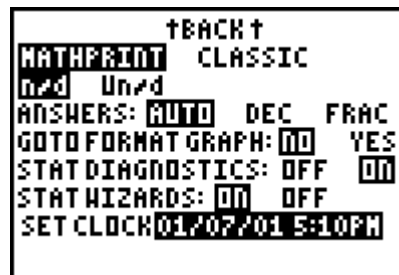
The Home screen can be used to answer questions 14 and 15 to predict the life expectancy and mortality rates for the numbers given.

Note: solve(expression, variable, guess).

The solve command is solving $Y_1(X) - 70 = 0$ for X.

If using Mathprint OS:

To access the function Y_1 quickly, press ALPHA [F4] and select **Y1**.
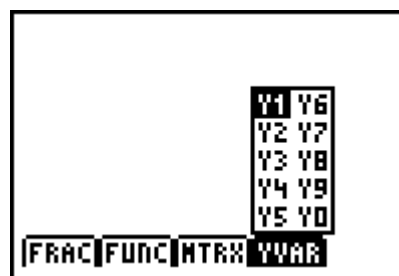
Solutions to Student Worksheet

	Heath Services	Safe Water	Sanitation
Mean Standard deviation	Mean = 70.15 sx = 25.01	Mean = 59.55 sx = 23.33	Mean = 49.35 sx = 27.61
Median Five-number summary	Min = 27 Q1 = 43 Median = 76 Q3 = 90 Max = 100	Min = 12 Q1 = 46 Median = 57 Q3 = 77.5 Max = 100	Min = 5 Q1 = 22.5 Median = 47.5 Q3 = 70 Max = 98

1. Heath Services: Best measure of central tendency is the median and the best measure of spread is the five-number summary because the data is skewed to the left.

    Safe Water: Best measure of central tendency is the mean and the best measure of spread is the standard deviation because the data is symmetric.

    Sanitation: Best measure of central tendency is the median and the best measure of spread is the five-number summary because the data is skewed to the right.

2. Sample response: Sanitation needs the greatest improvement since it has the lowest measure of central tendency. It also has the lowest minimum value. Over 50% of the countries have less than 50% of the people with access to sanitation. 75% of the countries have less than 70% of the people with access to sanitation.

3. Angola (46); Central Africa Republic (47); Zaire (52)

4. Students should justify any statement using a graph and/or a numerical statistic.
    Sample response: More money should be given to sanitation in these countries. This is the greatest need. This can be seen, as it has the lowest median value of the three variables measured.

Two Variable Statistics

5. Answers will vary. Life expectancy vs. safe water, sanitation, daily calorie intake and health all work. These variables give a positive correlation because an improvement in any of these variables should lead to a longer life. Daily calorie intake has the strongest relationship.

6. Answers will vary. Life expectancy vs. Infant mortality. Practically, as a country can reduce the death of infants, the life expectancy should increase.

7. Answers will vary. Safe water and health have very little relationship. Practically, countries with minimal finances will have to choose between health care and safe water for their people.

8. It appears linear with a negative correlation. It is a relatively strong relationship.

9. Answer will vary. Sample equation: $y = -0.2692x + 77.854$

10. $y = -0.207766x + 74.1402$

    The slope means that for each increase of .2 deaths per 1000 (or 2 per 10000) the life expectancy will decrease by 1 year. The units for slope are years per deaths per 1000.

    If there is an infant mortality rate of 0 then the life expectancy will be 74.14 years.

11. $r = -0.90797$

12. $r^2 = 0.824425$

13. It tells us that 82.44% of the variation in life expectancy is due to the infant mortality rate.

14. Life expectancy is 66.87 years.

15. The infant mortality rate is about 20.