



**Data**

Run the program **HEALTH**. Press **STAT** **ENTER** to see the data.

The data is a table of health indicators for some Asian and African countries in 1995. The numbers for infant mortality are the death rates among children under age 5 per 1000 live births. The numbers in the last three columns are the percentages of people who have access to health services, safe water, and sanitation.

© World Health Organization

**One Variable Statistics**

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

Make a histogram for each variable. Then make a series of three parallel box plots. On the Home screen, calculate the One Variable Statistics and record the information below.

	Health Services	Safe Water	Sanitation
Mean			
Standard deviation			
Median			
Five-number summary			

1. Describe which measure of central tendency and measure of spread describes each data set. Explain why the measures are appropriate.
2. Which of the three categories has the greatest need for overall improvement? Why?
3. Find a country that is in the bottom quartile for all three distributions. What quartile do they fall in for life expectancy?
4. You are asked to give your recommendations about the amount of money to spend in these areas. What would your response be? Be specific using graphs and numerical statistics to justify your position.



## Problem 2: Two-variable statistics

Investigate the relationships between the variables by constructing scatter plots. Find one with a positive correlation, one with a negative correlation, and one with no correlation.

- Describe your scatter plot with the **positive** correlation coefficient. What is the independent variable? Dependent variable? Practically, why does it make sense that these two variables give a positive correlation?
- Describe your scatter plot with the **negative** correlation coefficient. What is the independent variable? Dependent variable? Practically, why does it make sense that these two variables give a negative correlation?
- Can you find a scatter plot with relatively **no** correlation? Practically, why does it make sense that these two variables have no correlation?
- Now, graph **infant mortality versus life expectancy**. Describe the relationship between the two variables.
- Use **Manual-Fit** ( $\boxed{\text{STAT}}$  > CALC) to find your line of best fit.  
**Equation:** \_\_\_\_\_  
Explain what makes your line a line of best fit. Be specific.

10. Find the linear regression line. (Be sure to select **DiagnosticOn** from the Catalog first.)

**Equation:** \_\_\_\_\_

Interpret the slope of the equation. What does it tell you about the relationship between infant mortality and life expectancy? Be specific. Use correct units.

Interpret the  $y$ -intercept of the equation. What does it tell you? Be specific. Use units.

11. Find the correlation coefficient.  $r =$  \_\_\_\_\_.

12.  $r^2$  is called the **coefficient of determination**. It gives the percent of the variation in the dependent variable that can be explained by the linear relationship. Find  $r^2$ .

$r^2 =$  \_\_\_\_\_

13. What does the value of  $r^2$  specifically tell you about the relationship between infant mortality and life expectancy?

14. If the infant mortality rate of a country is 35, what is the life expectancy predicted to be?

15. If the life expectancy of a country is 70 years, what is the predicted mortality rate?