



Data

The data in the spreadsheet 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.

One Variable Statistics

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

To do so, create graphs and add measures of central tendency and spread. Describe why the measures of central tendency and spread that you chose are the most appropriate ones.

	Health Services	Safe Water	Sanitation
Best measure of central tendency			
Best measure of spread			
Justification			

1. Which of the three categories has the greatest need for overall improvement? Why?
2. Find a country that is in the bottom quartile for each of the three distributions. What country is it? What is their life expectancy?
3. 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.



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.

4. 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?
5. 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?
6. Can you find a scatter plot with relatively **no** correlation? Practically, why does it make sense that these two variables have no correlation?
7. Now, graph **infant mortality versus life expectancy**. Describe the relationship between the two variables.
8. Add a movable line and find your line of best fit.

Equation: _____

Explain what makes this movable line a line of best fit. Be specific.



9. Find the linear regression line.

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.

10. Find the correlation coefficient. $r =$ _____.

11. 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 .

$r^2 =$ _____

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

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

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