1. Scatter diagram : x = horizontal scale , y = vertical scale

2. Calculate the Linear Correlation Coefficient

   ρ = population linear correlation coefficient

   r = sample linear correlation coefficient (r estimates ρ)

   $$r = \frac{n\sum(xy) - (\sum x)(\sum y)}{\sqrt{n(\sum x^2) - (\sum x)^2}\sqrt{n(\sum y^2) - (\sum y)^2}}$$ (round off to 3 decimal places)

Test r for statistical significance

Method 1 : Hypothesis test

1. Write the claim about ρ.
   $\rho \neq 0$: "a correlation exists" (default claim)
   $\rho = 0$: "no correlation exists"
   $\rho > 0$: "a positive correlation exists"
   $\rho < 0$: "a negative correlation exists"
2. Write the opposite of the claim.
3. Identify H_0 and H_1 . (H_0 MUST CONTAIN EQUALITY)
4. Choose α and identify the test type. (look at H_1)
5. Find t_{crit} using Table A-3 Row = d.o.f. = n – 2 (bivariate)
   Column = α and test type
6. Calculate $t_{test} = \dfrac{r}{\sqrt{\dfrac{1-r^2}{n-2}}}$
7. Draw the bell curve to compare t_{test} and t_{crit} .
8. Reject H_0 if t_{test} falls in the critical region.
9. Write the conclusion.

Method 2 : Use Table A-6 on Page 780

1. Calculate r
2. On Table A-6 look at the row for n
3. If r > critical value, the results are statistically significant.

3. If r is significant, find the regression equation (Line of Best Fit) :

   for a population : $y = \beta_1 x + \beta_0$ for a sample : $\hat{y} = b_1 x + b_0$

   x = independent variable y = dependent variable

   b_1 = slope of the line (estimates β_1) = $\dfrac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2}$

   b_0 = y intercept (estimates β_0) = $\dfrac{(\sum y)(\sum x^2) - (\sum x)(\sum xy)}{n(\sum x^2) - (\sum x)^2}$

   round off b_1 and b_0 to 3 significant digits

4. Choose 2 different values for x between the minimum and maximum x values in your data.

5. Calculate $\hat{y}$ for the 2 values of x chosen using the regression equation.
   Result : Two new ordered pairs $(x_1, \hat{y}_1), (x_2, \hat{y}_2)$

6. Plot the 2 new ordered pairs on the scatter diagram and draw a line running through them.
   (this plots the regression line against the data)