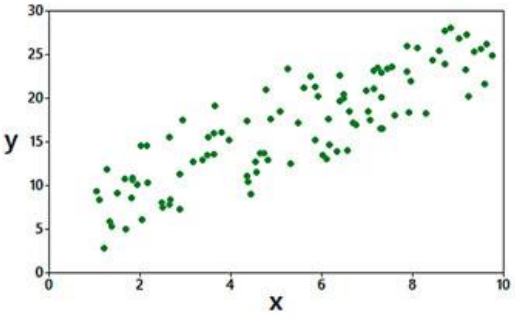
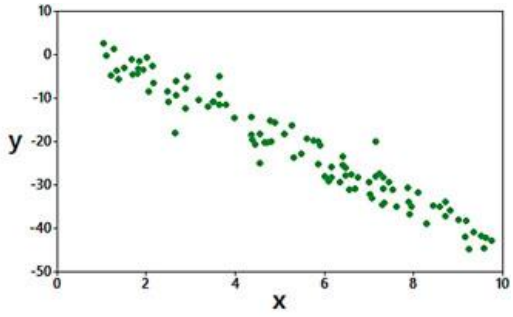
Correlation and Regression. Notes.

Correlation:

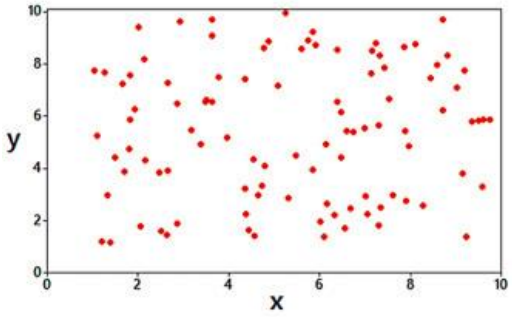
- A **correlation** exists between two variables when the values of one variable are somehow associated with the values of the other variable.
- A **linear correlation** exists between two variables when there is a correlation and the plotted points of paired data result in a pattern that can be approximated by a straight line.



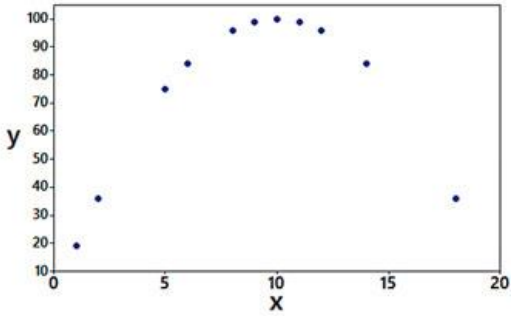
(a) Positive correlation: $r = 0.859$



(b) Negative correlation: $r = -0.971$



(c) No correlation: $r = 0.074$



(d) Nonlinear relationship: $r = 0.330$

We use the linear correlation coefficient *r*, which is a number that measures the strength of the linear association between the two variables.

Using *P*-Value from Technology to Interpret *r*: Use the *P*-value and significance level α as follows:

P-value $\leq \alpha$: Supports the claim of a linear correlation.

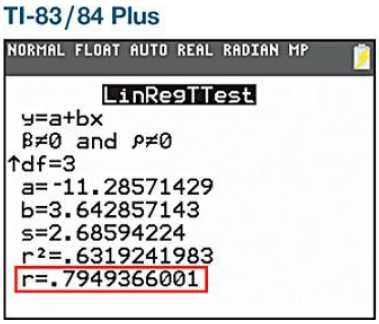
P-value $> \alpha$: Does not support the claim of a linear correlation.

Using Table A-5 to Interpret *r*: Consider critical values from Table A-5 or technology as being both positive and negative:

- **Correlation** If $|r| \geq$ critical value, conclude that there is sufficient evidence to support the claim of a linear correlation.
- **No Correlation** If $|r| <$ critical value, conclude that there is not sufficient evidence to support the claim of a linear correlation.

TI83 & TI84:

Enter the x or independent variable on L1, and y or dependent variable on L2.
Press STAT, TESTS, choose, LinRegT Test:



Statdisk.org:

Analysis, Correlation and Regression, update significance, enter x variable on column 1; y in column 2, Select the columns to be used for the x and y variables on the left menu and hit Evaluate.

Example:

```
Sample Size, n:   6
Degrees of Freedom: 4

Correlation Results:
Correlation Coeff, r:  0.41079
Critical r:            ±0.8114
P-Value (two-tailed):  0.41847

Regression Results:
Y= b0 + b1x:
Y Intercept, b0:      0.95
Slope, b1:            0.225
```

Therefore, the equation is:
y = 0.95 + 0.225 x

In this example the equation is meaningless since the p value is larger than alpha; therefore, there is no evidence of correlation between the two variables. The best predicted value for a given value of x would be the mean value of y.
In case where there is a significant correlation between the variables, the best predicted value for a given value of x is obtained by substituting the x-value into the regression equation.