Correlation Analysis

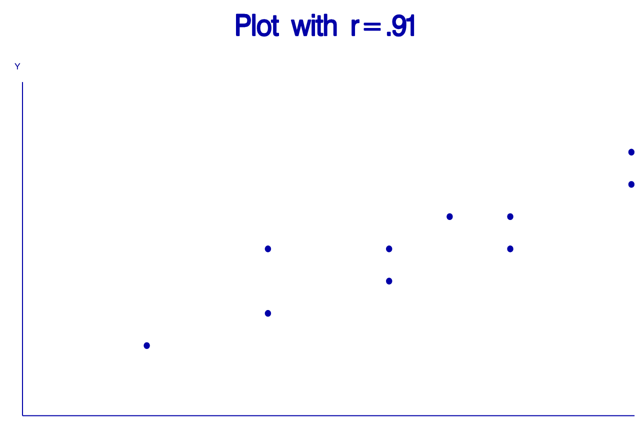
Important Terms and Definitions

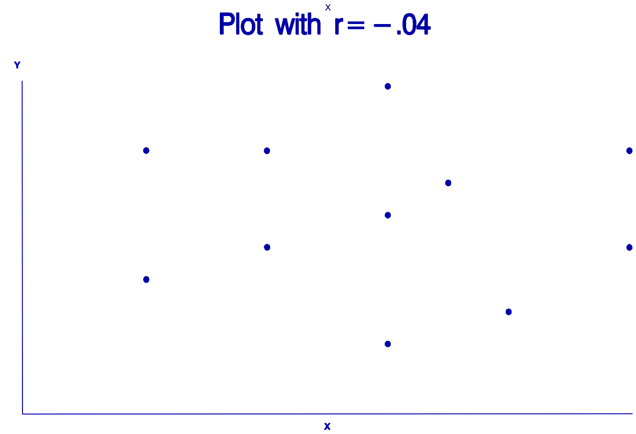
Correlation analysis is a method used to determine the strength of the linear relationship between two variables (X and Y).

The parameter which measures the strength of the linear relationship is called **rho (ρ)**.

Properties of ρ
1. If ρ=0, then the variables X and Y are **uncorrelated**.
2. If ρ>0, then we say there is a **positive correlation** between X and Y. A positive correlation indicates that as the X values increase, so do the Y values.
3. If ρ<0, then we say there is a **negative correlation** between X and Y. A negative correlation indicates that as the X values increase, the Y values will decrease.
4. The linear correlation parameter ρ are between -1 and 1, inclusive.
5. If ρ=1 or -1, then we say there is a **perfect linear correlation** between variables X and Y.

One estimator of the linear correlation (ρ) is called the **Pearson product moment coefficient of correlation** (r).





Formula for the Pearson Product Moment Coefficient of Correlation

$$r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$$

where

$$S_{xx} = \sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n} \qquad S_{xy} = \sum_{i=1}^{n} x_i y_i - \frac{\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n}$$

$$S_{yy} = \sum_{i=1}^{n} y_i^2 - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}$$

A Hypothesis Test for the Correlation Parameter

Hypotheses

H_o: $\rho = 0$	H_o: $\rho = 0$	H_o: $\rho = 0$
H_a: $\rho < 0$	H_a: $\rho > 0$	H_a: $\rho \neq 0$

Test Statistic

$$t = r \cdot \sqrt{\frac{n - 2}{1 - r^2}}$$

Distribution for the Rejection Region

The rejection region is found using a t-distribution with df = n-2.

Example: In the table below is repeated the random sample of rents taken from a rental pricing survey. All of the apartments have air conditioning, are unfurnished and have no dishwasher.

Apartment	Number of Bedrooms (X)	Monthly Rent (Y)	X^2	Y^2	XY
Charleston Ave.	1	300	1	90000	300
Clarendon Drive	1	305	1	93025	305
College Street Apartments	1	325	1	105625	325
Brookdale Apartments	2	375	4	140625	750
Porter House Apartments	2	400	4	160000	800
Robin Hill Apartments	2	425	4	180625	850
Robin Hill Apartments	3	500	9	250000	1500
Berkley Drive	3	550	9	302500	1650
Sum	15	3180	33	1322400	6480

Would you expect the coefficient of correlation to be positive or negative? Why?

Calculate the Pearson product moment coefficient of correlation between the number of bedrooms and the monthly rent. Is the coefficient positive or negative? Is that what you expected to happen?

Test if the linear correlation is significantly different from zero ($\alpha = .05$)

Example: Metal fatigue is the structural damage that results when a material is subjected to fluctuating stresses and strains. Engineers are often interested in determining the fatigue life (cycles to failure) of a device which is subjected to a specific fluctuating stress. Below are test data on the fatigue life of steel.

Fatigue Life of Steel			
Stress (X)	Life (Y)	Stress (X)	Life (Y)
11.5	3280	40.1	750
14.3	1563	40.1	316
16.0	977	47.3	242
24.0	804	60.5	283
24.0	482	67.0	187
34.5	736	75.4	200

Below is a correlation matrix of the variables and log transformations of the variables.

Correlation Matrix				
	Stress	Life	Log(Stress)	Log(Life)
Stress	1.00000	-0.68884	0.96802	-0.87740
Life	-0.68884	1.00000	-0.80327	0.90387
Log(Stress)	0.96802	-0.80327	1.00000	-0.92097
Log(Life)	-0.87740	0.90387	-0.92097	1.00000

If you wanted to predict the life (based on stress) using a linear relationship, which variables should be used?

a. Stress vs. Life b. Stress vs. Log(Life) c. Log(Stress) vs. Life d. Log(Stress) vs. Log(Life)

Plots for the Fatigue Life Example