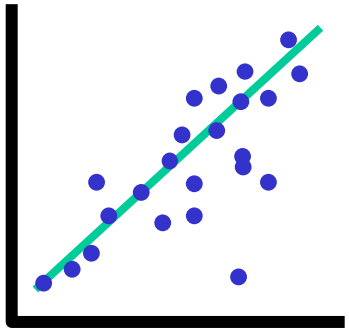
Continuous Improvement Toolkit

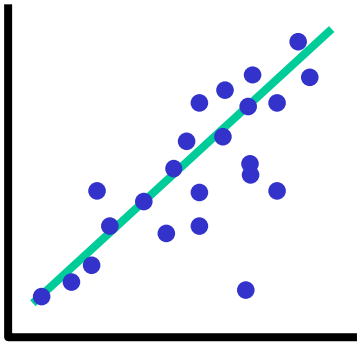
Correlation




Managing Risk
PDPC
FMEA
RAID Logs
Fault Tree Analysis
Risk Assessment*
Traffic Light Assessment
Deciding & Selecting
Pros and Cons
Importance-Urgency Mapping
Break-even Analysis
Cost -Benefit Analysis
Force Field Analysis
Voting
Pugh Matrix
Decision Tree
TPN Analysis
QFD
Matrix Diagram
Kano Analysis
Prioritization Matrix
Critical-to Tree
Paired Comparison
Planning & Project Management*
RACI Matrix
Stakeholders Analysis
PEST
PERT/CPM
Activity Diagram
SWOT
Roadmaps
Project Charter
Gantt Chart
PDCA
Control Planning
Gap Analysis
Hoshin Kanri
Kaizen
How-How Diagram
Tree Diagram**
Standard work
Lean Measures
KPIs
OEE
Capability Indices
MSA
RTY
Descriptive Statistics
Cost of Quality
Reliability Analysis
Graphical Analysis
Cause & Effect Matrix
Pareto Analysis
Confidence Intervals
Understanding Cause & Effect
Probability Distributions
ANOVA
Hypothesis Testing
Design of Experiments
Scatter Plot
Correlation
Regression
Multi-Vari Charts
5 Whys
Chi-Square Test
Relations Mapping*
Identifying & Implementing Solutions***
Simulation
TPM
Mistake Proofing
Pull Systems
JIT
Ergonomics
Work Balancing
Automation
Bottleneck Analysis
Visual Management
Flow
Value Analysis
5S
Understanding Performance
Run Charts
Control Charts
Benchmarking
Sampling
Fishbone Diagram
TRIZ***
Brainstorming
Analogy
SCAMPER***
Nominal Group Technique
Mind Mapping*
Affinity Diagram
Attribute Analysis
Lateral Thinking
Visioning
Creating Ideas**
Wastes Analysis
SMED
Time Value Map
Process Redesign
IDEF0
Value Stream Mapping
SIPOC
Flow Process Chart
Process Mapping
Flowcharting
Service Blueprints
Designing & Analyzing Processes
Data Collection
Focus groups
Interviews
Photography
Check Sheets
Measles Charts
Surveys
Critical Incident Technique
Observations

- Correlation

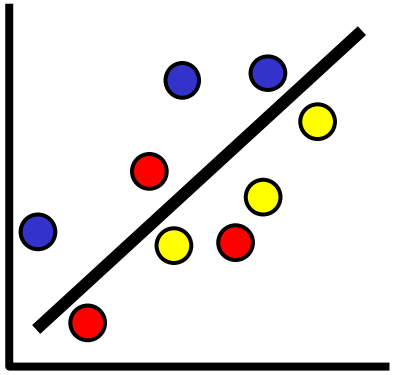
- **Correlation** (& Regression) is used when we have data inputs and we wish to explore if there is a relationship between the inputs and the output.
  - What is the strength of the relationship?
  - Does the output increase or decrease as we increase the input value?
  - What is the mathematical model that defines the relationship?



- Given multiple inputs, we can determine which inputs have the biggest impact on the output.
- Once we have a model (regression equation) we can **predict** what the output will be if we set our input(s) at specific values.

- Correlation

- Correlation is the degree to which two continuous variables are related and change together.
- It is a measure of the **strength** and **direction** of the linear association between two quantitative variables.
- Uses the Scatter Plot representation.



- Correlation

Example:

- A market research analyst is interested in finding out if there is a relationship between the sales and shelf space used to display a brand item.
- He conducted a study and collected data from 12 different stores selling this item.
- **Practical Problem:**
  - Is there a relationship between sales of an item and the shelf space used to display that item?
  - If there is a relationship, how strong is it?
- **Statistical Problem:**
  - Are the variables ‘Sales’ and ‘Shelf Space’ correlated?

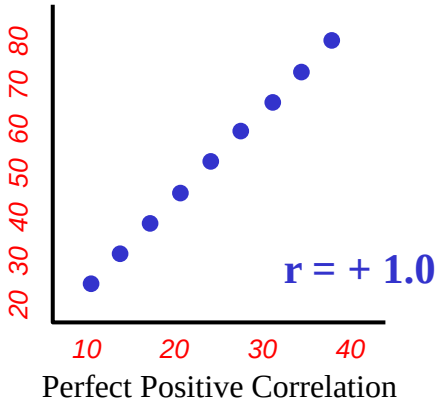
- Correlation

Other Examples:

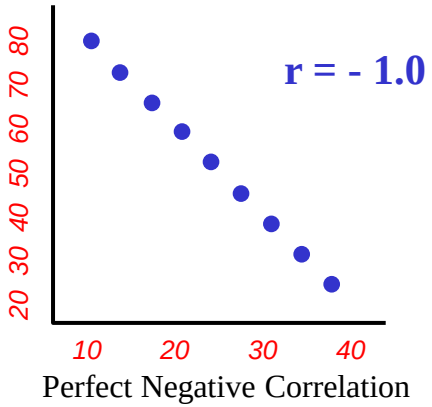
- The relationship between the height and the width of the man.
- The relation of the number of years of education someone has and that person's income.
- The relationship between the training frequency and the line efficiency.
- The relationship between the downtime of a machine and its cost of maintenance.

- Correlation

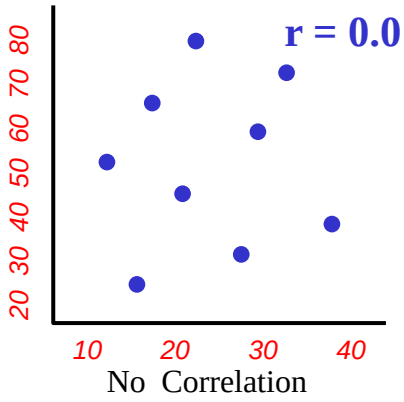
- Correlation coefficient or **Pearson's correlation coefficient (r)** is a way of measuring the strength and direction of linear association.
- The coefficient ranges from +1 (a strong direct correlation) to zero (no correlation) to -1 (a strong inverse correlation).



Perfect Positive Correlation



Perfect Negative Correlation

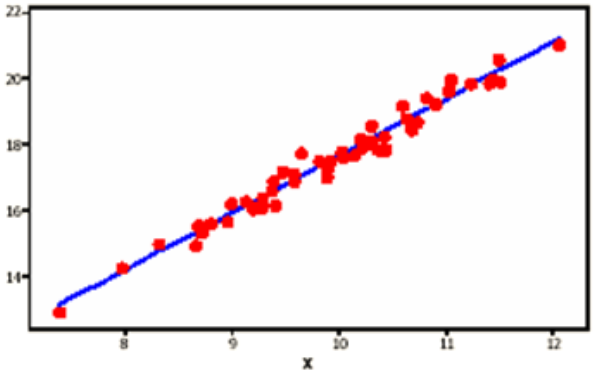


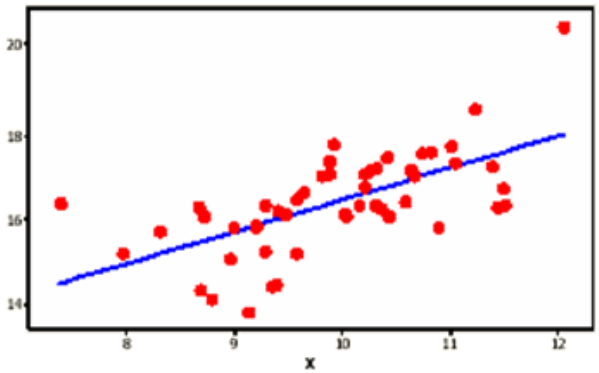
No Correlation

- Correlation

Example - The Strength and Direction of Linear Association:

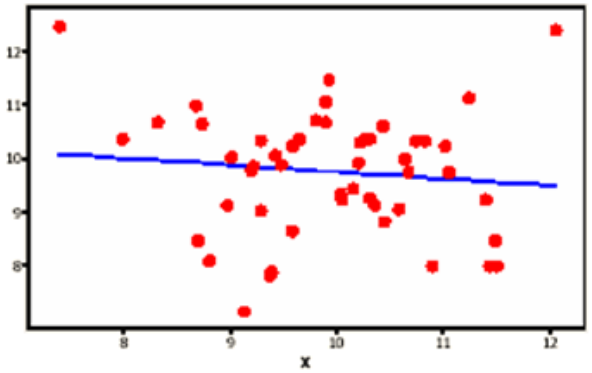
Strong Positive
r = 0.986

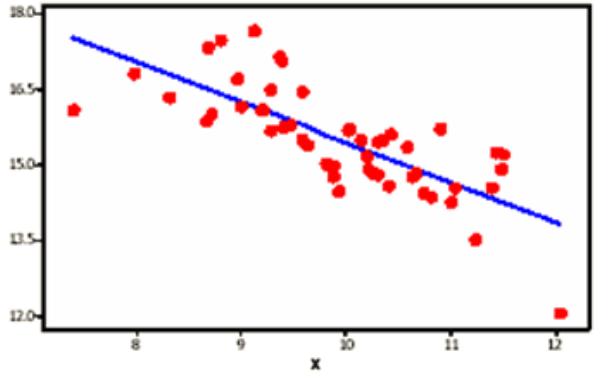




Moderate Positive
r = 0.641

Weak Negative
r = -0.111



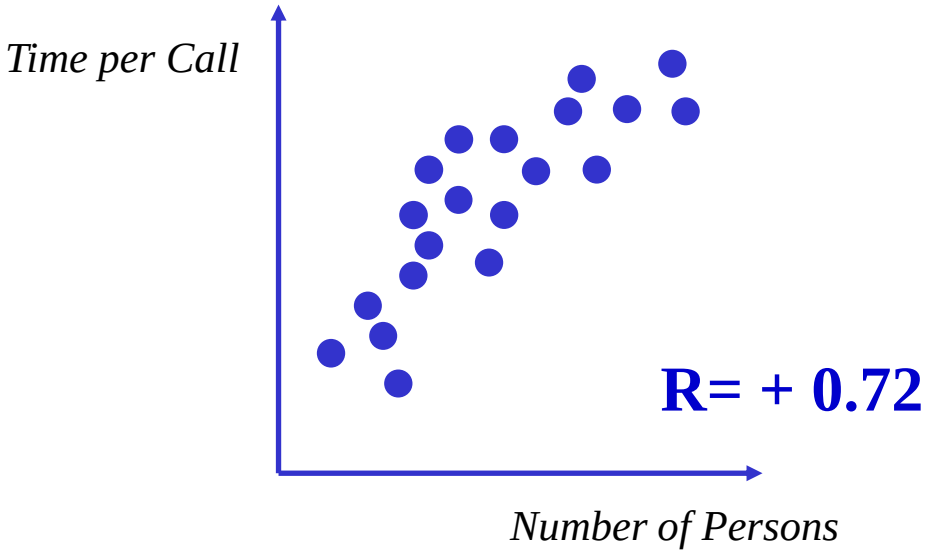


Moderate Negative
r = -0.755

- Correlation

Example – The Number of Personnel and the Time per Call:

- Is there is a correlation?

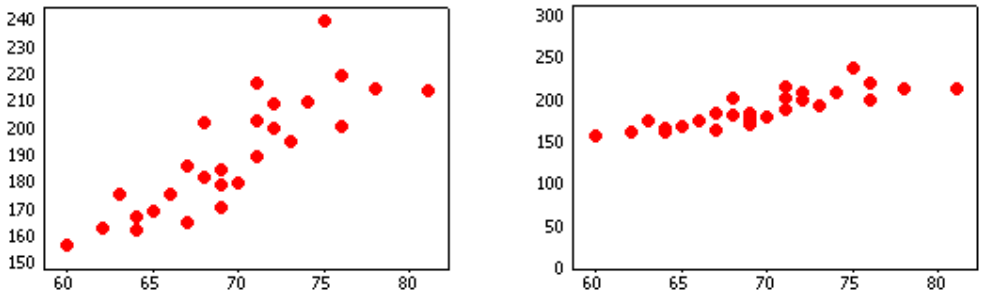


Answer:

- There is a direct (positive) relationship.
- It suggests that the more personnel the longer they spend on each call.

- Correlation

- Can we relay on the scatter plot on finding the relationship between the variables?
- **Questions:** Which data have stronger relationship in the following scatter plots?



Answer: Both graphs plot the same data (the ranges are different), their correlation coefficients are the same.

- Correlation

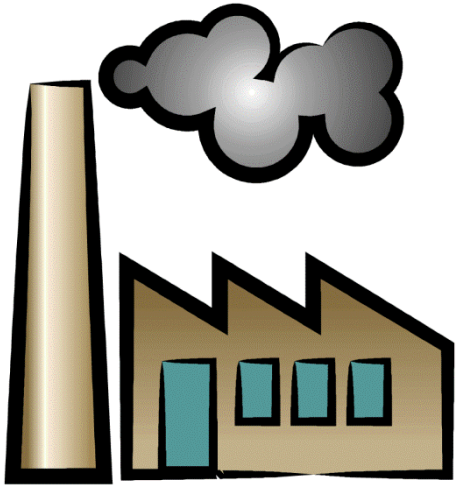
Hints:

- Because of the random nature of data, it is possible for a scatter plot (or the Pearson coefficient) to suggest a correlation between two factors when in fact none exists.
- This can happen where the scatter plot is based on a small sample size.
- The statistical significance of your Pearson coefficient must be assessed before you can use it.
- Correlation does not imply causation!
- Always think which factor is the real “cause”.
- Two things exist together but one does not necessarily cause the other.

- Correlation

Coincidence:

- Since the 1950s, both the atmospheric CO2 level and crime levels have increased sharply.
- Atmospheric CO2 causes crime.
- The two events have no relationship to each other.
- They only occurred at the same time.



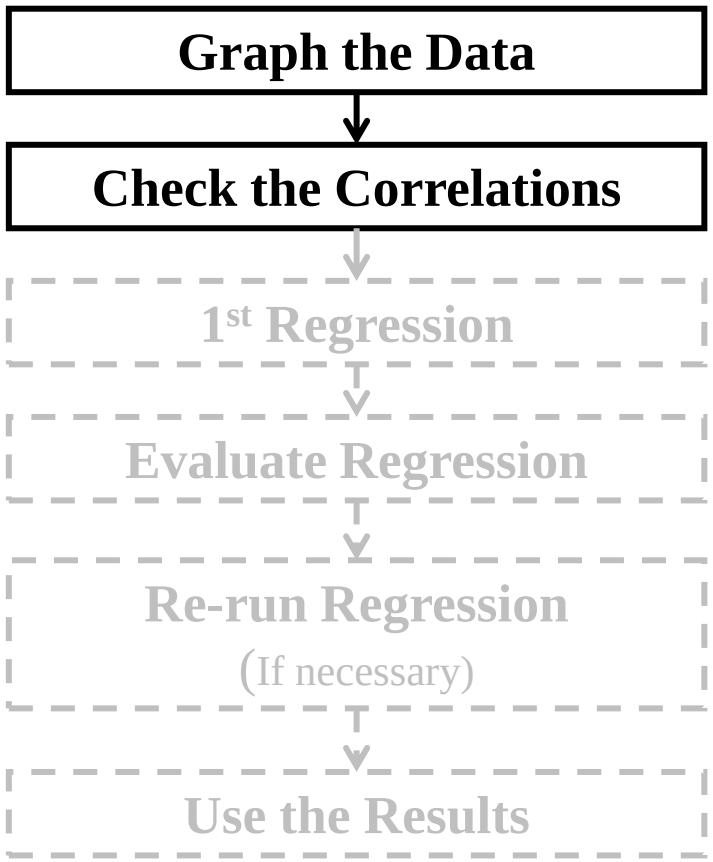
- Correlation

Hidden Factors:

- In London a survey pointed out a correlation between accidents and wearing coats (taxi drivers).
- It was assumed that coats could hinder movements of drivers and be the cause of accident.
- A new law was prepared to prohibit drivers to wear coats when driving.
- Finally another study pointed out that people wear coats when it rains! Rain was the hidden factor common to wearing coat and accident frequency.

- Correlation

The Process:



Scatter plot

Use Pearson Coefficient

Linear / Multiple regression

R-squared & analyze residuals

Simple: With different model (Cubic)
Multiple: Remove unnecessary items

Control critical process inputs & select best operating levels.