TOOL 18

Scatter plots

Overview

Scatter plots (also known as scatter diagrams or scattergrams) are used to study possible relationships between two variables. The purpose of the scatter plot is to display what happens to one variable when another variable is changed. The scatter plot is used to test a theory that the two variables are related. This can help ensure that you are focusing improvement efforts on the true cause of a problem.

Scatter plots can be used to explore:

- If there is a relationship between outside temperature and cases of the common cold? As temperatures drop, does the number of colds increase?
- Is there a relationship between length of time spent with a doctor and patient satisfaction rates?
- If there is a relationship between number of patients in an Emergency Department and length of wait-time to see a doctor?

Structure of a scatter plot

A scatter plot is composed of a horizontal axis containing the measurements of one variable (independent variable) and a vertical axis representing the measurements of the second variable (dependent variable).

- **X- axis** (horizontal axis): measurements of the independent variable. This variable is the one that you predict may affect the other. For example, age.
- **Y-Axis** (vertical axis): measurements of the dependent variable. This is variable is affected by other factors, and can (potentially) be influenced by the independent variable. For examples, falls.

If you are not sure which variable is which, you can plot them on either axis.

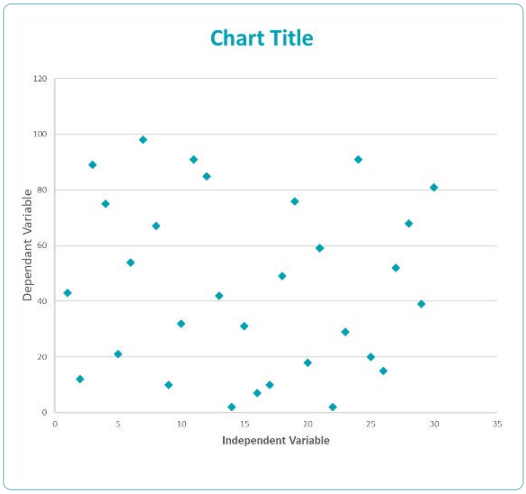


Figure 1: Scatter plot structure

How to create a scatter plot

Scatter plots can be created in many ways, the simplest is to simply draw one on a white board or on paper but they can also be easily created in Excel. Scatterplots can be started with only one data point but they work best with at least 30 observations.

1. Start with a hypothesis, for example:
   - As X (independent variable) increases, do you think Y (dependent variable) will also increase?
   - If X increases will Y decrease?
   - Is there any relationship between X and Y?
2. Collect your data: the data should be collected and recorded in pairs. For example, for each patient has their wait time in minutes 'paired' with their level of satisfaction.
3. Plot the data points or add data to the scatter plot template.
4. Interpret the patterns on the chart which indicate the type of relationship between the variables.

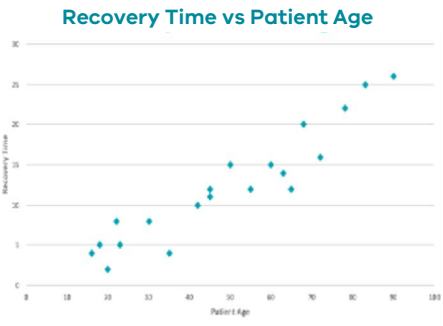
Table 1: Strength of Pearson's r values

Size of Correlation of Pearson's r	Interpretation
.9 to 1	Very Strong
.7 to .9	Strong
.5 to .7	Moderate
.3 to .5	Weak
0 to .3	No Relationship

Important considerations

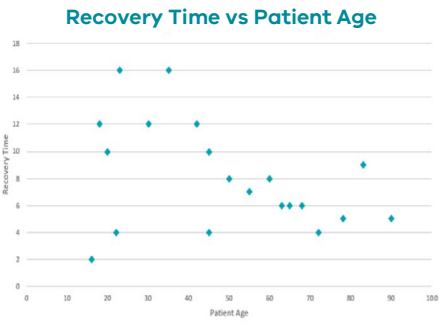
A cautionary note: a scatter plot shows patterns in data and indicate the existence of a relationship between two variables. However, the graph only cannot confirm for sure that there is a direct cause-and-effect relationship between the two variables, as there may be other factors that affect the variables tested.

Scatter plot examples: Relationship between recovery time and patient age



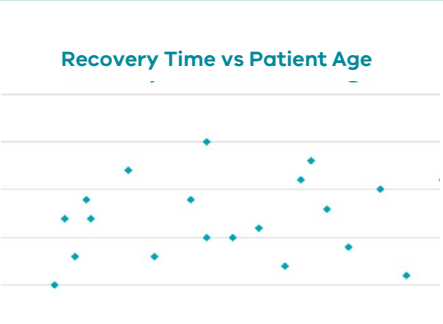
Strong Pearson's r = 0.9
Positive
Linear

Recovery times are likely to increase as age increases



Weak Pearson's r = - 0.4
Negative
Linear

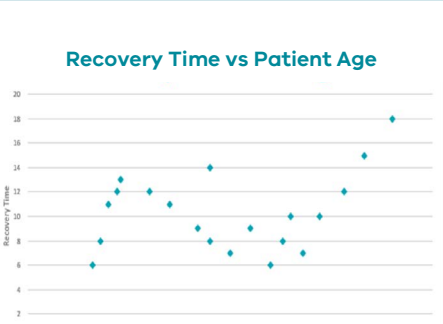
Other variables in addition to age may affect recovery time



Pearson's r = 0.1

No relationship

There is no demonstrated connection between age and recovery time



Pearson's r = 0.17

Non-Linear

Initial increase in recovery time with age but once above 20, recovery time decreases, reaching a minimum, and then increases again as age continues to advance.

Additional resources

To learn more about Quality Improvement you can access the following resources:

- SCV Quality Improvement Toolkit
- Institute for Healthcare Improvement website
- NSW Clinical Excellence Commission Quality Improvement Tools

Content adapted with permission from the Institute for Healthcare Improvement (IHI) and the Clinical Excellence Commission (CEC)