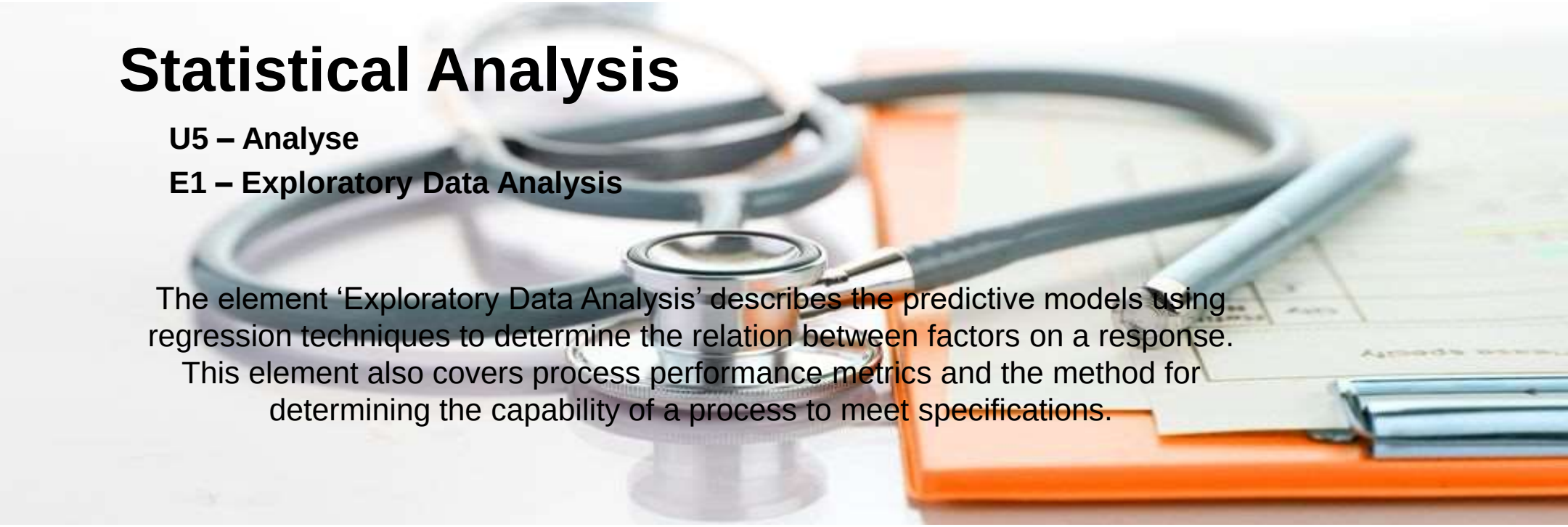
LSSH
lean six sigma for healthcare

Statistical Analysis

U5 – Analyse

E1 – Exploratory Data Analysis



The element ‘Exploratory Data Analysis’ describes the predictive models using regression techniques to determine the relation between factors on a response. This element also covers process performance metrics and the method for determining the capability of a process to meet specifications.

Lifelong Learning Programme

Yellow Belt

Correlation analysis

Correlation studies the degree of correlation between two variables

Correlation does not mean that there is a cause and effect relation

Example:

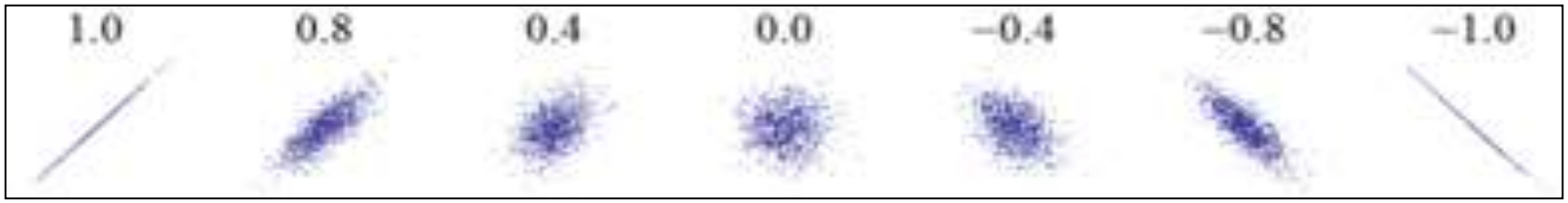
- **Is there a correlation between the age of the patient and efficiency of the administred drug?**

Statistical Analysis

Correlation coefficient

- **The Pearson correlation coefficient is used to measure the strength of the linear relationship between two variables**

- **The correlation coefficient assumes a value between -1 and +1**



Statistical Analysis

Correlation coefficient

The correlation coefficient (R) lies between -1 and +1

- **‘-1’ depicts complete inverse (negative) dependence**
- **‘0’ depicts complete independence**
- **‘+1’ depicts complete direct (positive) dependence**

General Rules

- **Strong : correlation coefficient |R| > 0.8**
- **Moderate : correlation coefficient 0.5 < |R| < 0.8**
- **Weak : correlation coefficient |R| < 0.5**

Regression Analysis

Regression analysis investigates the relationship between cause and result

Examples:

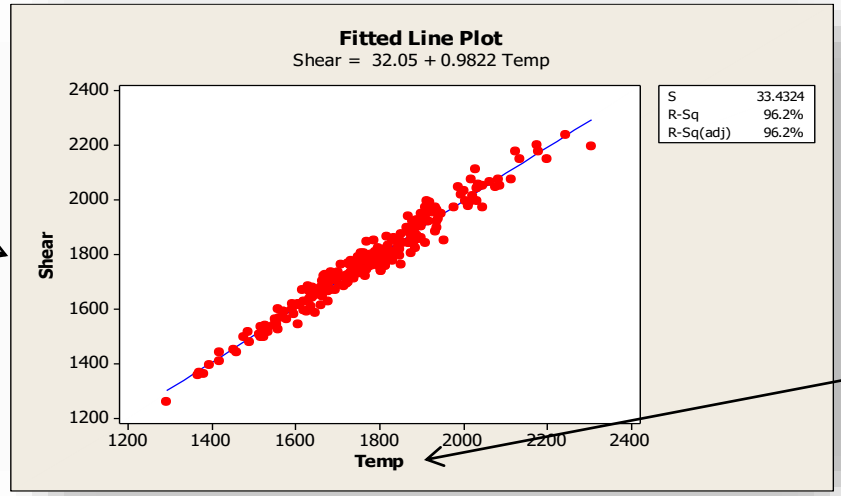
- **When temperature increases, the process runs faster**
- **As you get older, more accidents happen**
- **The warmer it gets, the more ice-cream people eat**

Statistical Analysis

Factor and response

- **The ‘Factor’ is the independent variable X**
- **The ‘Response‘ is the value that changes as a result of a changing ‘Factor’**
- **The ‘Response’ is the dependent variable Y**

Response Y
(Effect)



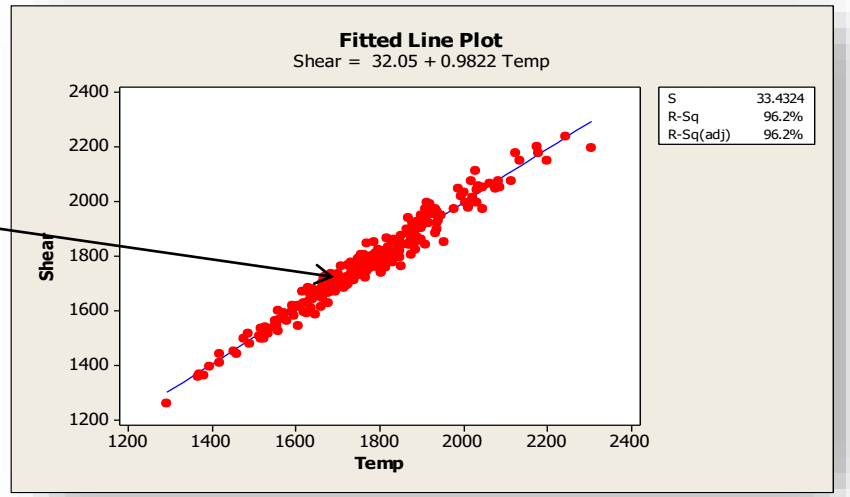
Factor X
(Cause)

Mathematical context

- **Regression analysis indicates the relation between the dependent Response (Y) and the independent Factor (X)**

- **Singular Linear Regression investigates the relation between one continuous Y and one continuous X**

Regression line



LSSH
lean six sigma for healthcare

Hypothesis Testing

U5 – Analyse

E2 – Hypothesis Testing

The element ‘Hypothesis testing’ reviews test methods that are used to test a hypothesis. This element also discusses Confidence Intervals that indicate the reliability of test conclusions.

Lifelong Learning Programme

Yellow Belt

Hypothesis Testing

Hypothesis

A hypothesis is a statement that something is true:

- **Based on this hypothesis, we predict the expected outcomes of the test**
- **If the outcome of the test has a low probability (unlikely), we will reject the hypothesis**
- **However, there will always be a chance that we reject a true hypothesis**

Hypothesis testing

In processes, we test hypotheses in the same way:

- **We do not want to react to common cause variation**
- **We only want to react to uncommon (special cause) variation**

In our example of the Elderly home centre:

- **We do not want to replace the stove when the variation was caused by not efficient walking routes**

- **We do want to replace the stove when this was the cause of the high variation in the temperature of the meals**

Hypothesis testing

‘One is presumed innocent until proven guilty’

This also applies to hypothesis testing:

- **The null-hypothesis (H_0) always assumes there is no difference**
  Even when we suspect that there actually is a difference!

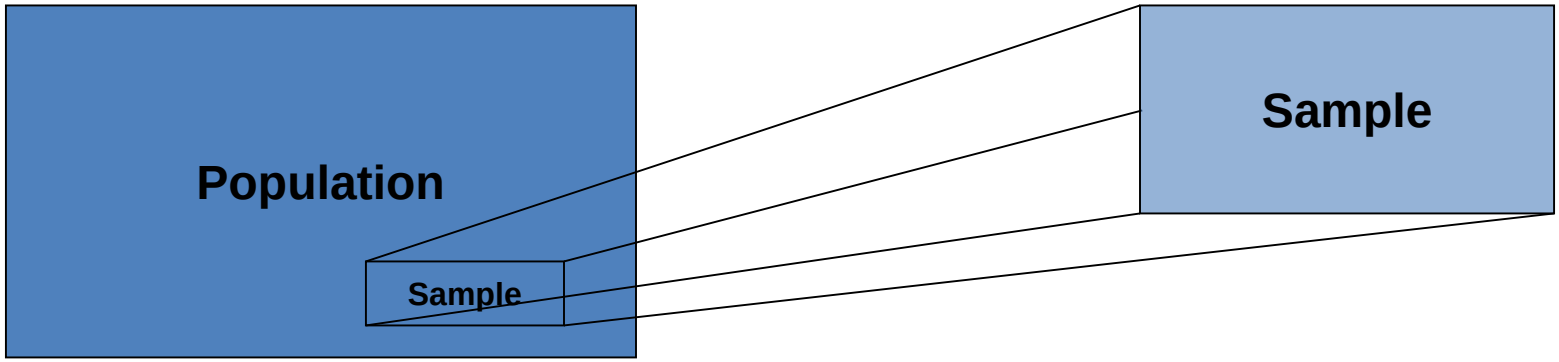
- **The alternative hypothesis (H_a) describes the difference**
  This is an assumption that must be reviewed and proven

Hypothesis Testing

Benefits of hypothesis testing

- **Helps to carefully handle uncertainties**
- **Prevents subjective interpretations**
- **Helps when making risky decisions**
- **Statistically quantifies the uncertainty**

Hypothesis Testing



Population parameters

μ = Population mean
σ = Population standard deviation

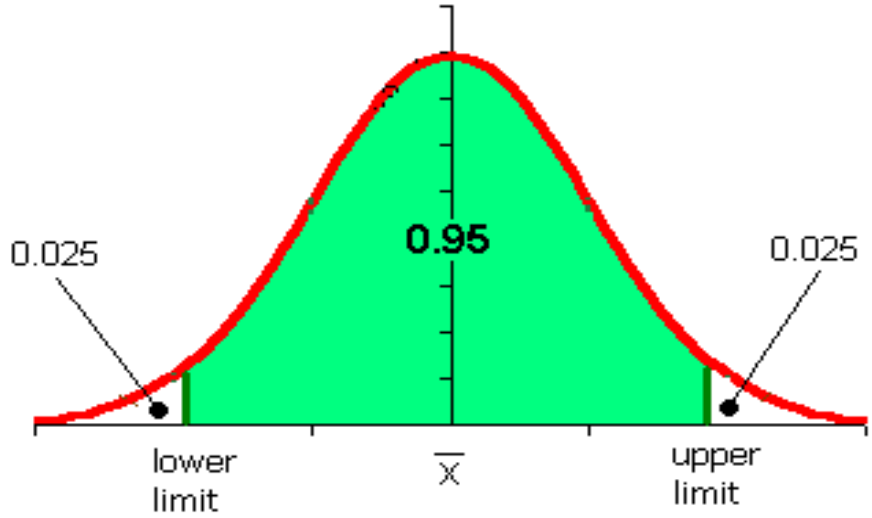
Sample statistics

$\bar{x}$ = Sample mean
s = Sample standard deviation

Hypothesis Testing

Confidence interval

- Statistics such as the average ($\bar{X}$) and standard deviation (s) of the sample are only estimators, not the real values!
- From sample to sample these estimates will differ
- A so-called confidence interval indicates how reliable the estimate is
- 95% is often used as confidence level



Hypothesis Testing

Influence of the sample size