Anexas Consultancy Services

Module 6: Health Data Analytics

Topics Covered

- Assist in evaluating and developing data management systems to support quality improvement
- Design data collection plans
- Identify and select measure
- Identify external data sources for comparison and benchmarking
- Design scorecards and dashboards for different audiences+ Compare data sources to establish benchmarks
- Collect and validate quantitative and qualitative data
- Use data management systems for organization, analysis and reporting of data
- Use statistics to describe data and examine relationships
- Use data visualization and display techniques
- Use statistical process control techniques and tools
- Use measurement tools to evaluate process improvement

Anexas Consultancy Services

Assist in evaluating and developing data management systems to support quality improvement

Types of Information Systems

- Administrative
- Clinical
- Decision

Types of Information Systems

- Administrative support information systems
- Aid day-to-day operations in healthcare organizations.
- Examples: This includes financial information systems including payroll, accounts payable, patient accounting, cost accounting, forecasting, budgeting ; human resource systems with employee records , position and performance management, labour analysis , turnover and absenteeism; and office automation systems with word processing, email, scheduling, facsimile/scanning, electronic spreadsheets.

Types of Information Systems

- Clinical information systems are designed to support direct patient care processes. Automated clinical information systems have great potential for analysing and improving quality of patient care.
- Examples: EHRs and their retrieval system; computer-assisted medical decision-making for histories and physicals and antibiotic selection; clinical application programs for health-risk programs, health maintenance organization encounter data, clinical algorithms, predictive modelling , and utilization.

Types of Information Systems

- Decision support systems address strategic planning functions.
- Decision support data facilitate cross-functional analysis to improve patient care processes and outcomes.
- Examples: strategic planning and marketing , resource allocation, performance evaluation and monitoring, product evaluation and services, and medical management (e.g., Evidence-based Practice or EBP, clinical guidelines, and clinical pathways).

Decision Support

- Helps in making comparison with competitors
- Identifies practitioners and providers who System meet acceptable levels of quality
- Allows providers to respond rapidly to market changes Justifies pay for exceptional performance
- Analyses and interprets outcomes data
- Used to develop outcomes information management plan

Decision Support System

- Strategic planning and marketing
- Resource allocation
- Performance evaluation and monitoring
- Product evaluation and services
- Medical management
  - Evidence-based practice
  - Clinical guidelines
  - Clinical pathways

Medical Record Analysis

- Chart (EHR) based
  - Nursing/medical record analysts review individual records
  - Nursing/medical record analysts input diagnostic procedural, and detailed clinical findings
  - System has higher cost and smaller sample size
  - Can be used to identify severity-adjusted and risk-adjusted information
  - Severity-adjusted: Measurement data
  - Risk-adjusted: Count data

Medical Record Analysis

- Code-based
  - Based on retrospective administrative data such as Uniform bill document (UB-92)
  - Uses clinical information spanning entire stay
  - Has lower cost and larger sample size
  - Submission of payer data deemed public information required by states

Anexas Consultancy Services

Design data collection plans

Data versus Information

DATA

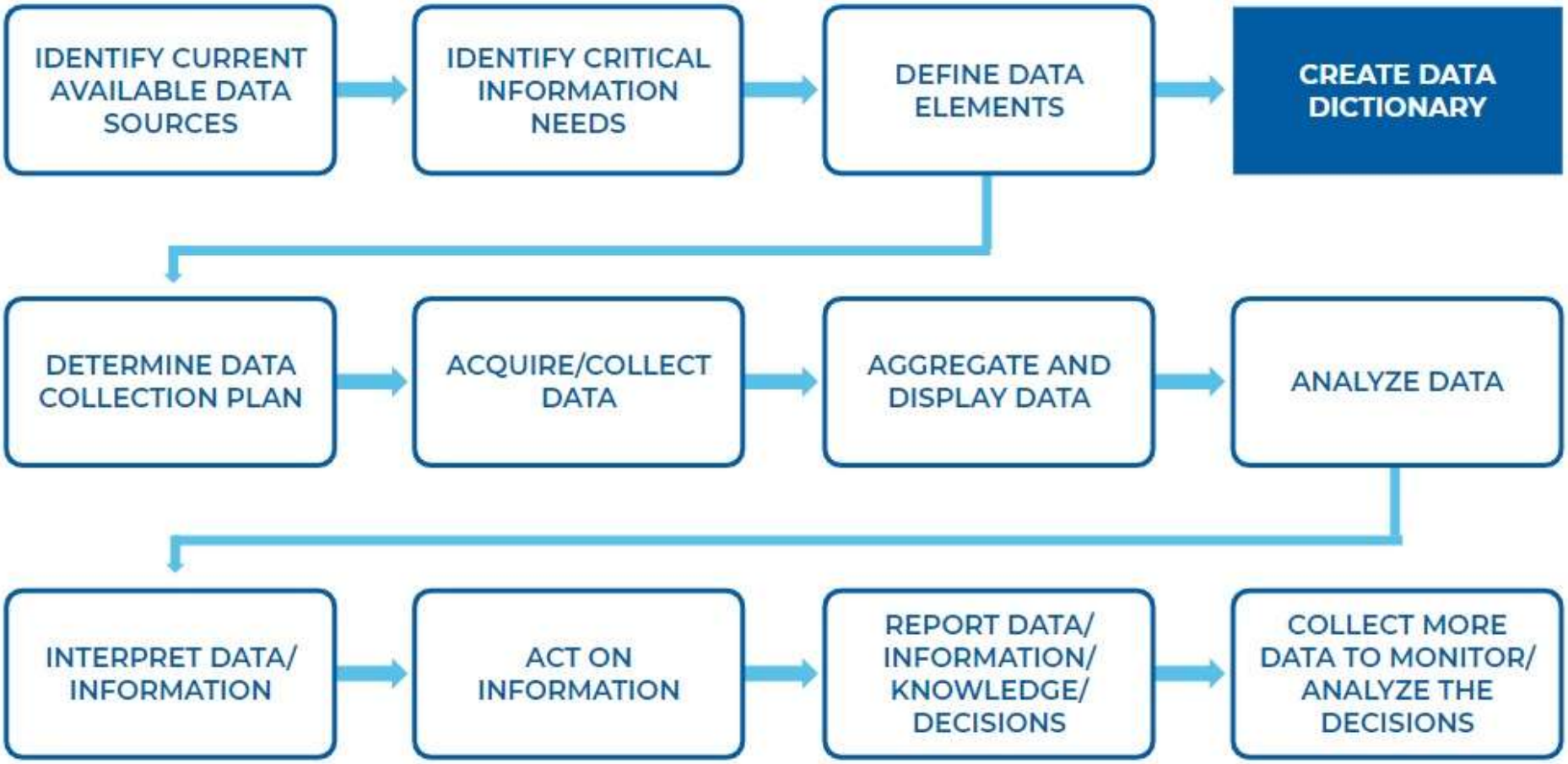
- Abstract representations of things, facts, concepts, instructions that are stored in a defined format and structure.

INFORMATION

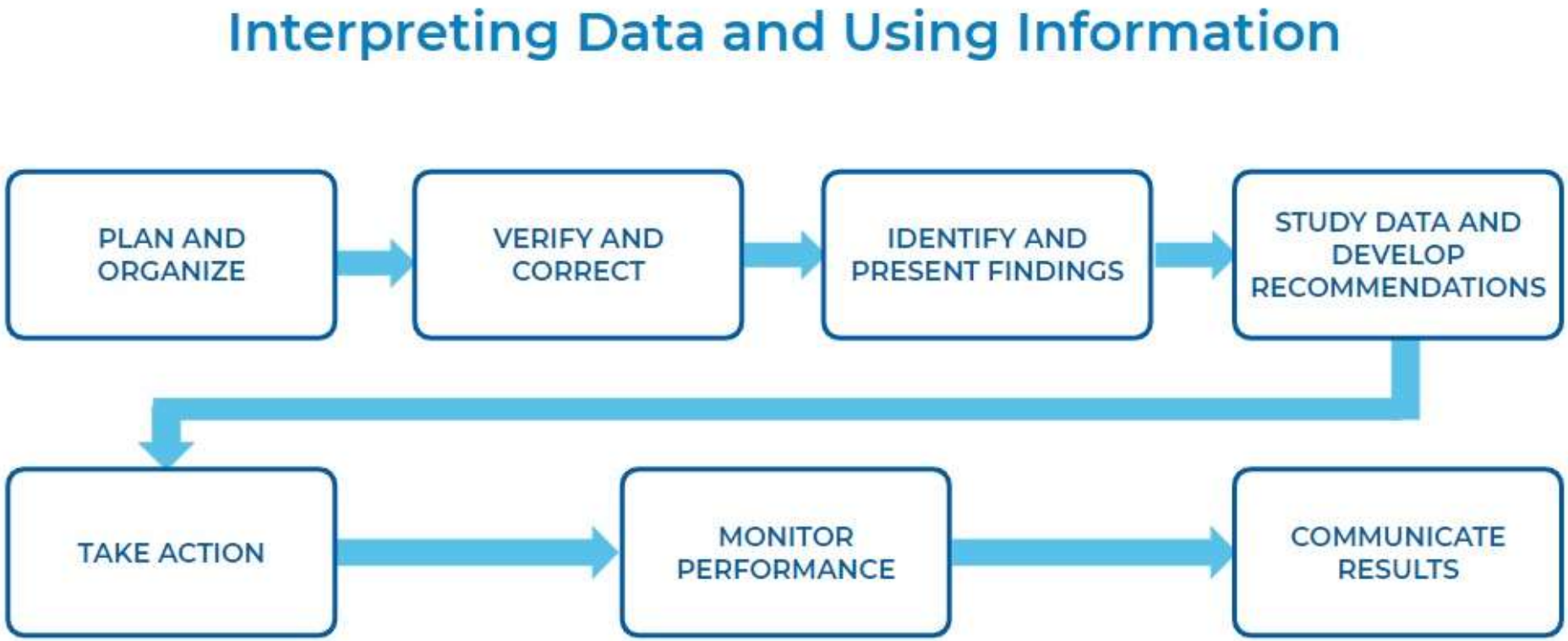
- Is obtained when data is translated into results and statements that are useful for decision making

DATA → INFORMATION →DECISIONS

Data Collection Process



Interpreting Data and using Information



Analysing Data

Report

- Report and analyse data regularly-
- Consider timelines of internally gathered data, internal data gathered by external sources and external data

Validate

- Validate accurate data collection

Display

- Use graphs to display data and include a table of values- Display data in easily understood format
- Provide a brief summary
- Provide contextual background

Analysing Data

Explain

- Explain data collection specifics (how, when, where, from whom)

Summarize

- Summarize meaning of values and how they were computed- Identify removed outliers+ Include time order of the graph

Analyze

- Analyze random variances and identify unexpected patterns- Example: Common Cause variation, Special cause variation, Non-random signal of change

Sampling designs

Basic Sampling Designs

- Population (N): total aggregate or group
- Sample (n): a portion of the population
- Sampling:

  Provides a logical way of making statements about a larger group« Allows quality professionals to make statements or generalize from the sample to the population depending on the type of sampling used

Sampling designs

- Probability Sampling

  Every element in the population has an equal or random chance of being selected.

- Nonprobability Sampling

  It is not possible to estimate the probability that every element has been included.

Sampling designs

Types of Probability Sampling

- Simple Random Sampling

  Each individual in the population has an equal chance to be chosen.

  Example: Put all names in a hat; pull one for a door prize.

- Systematic Sampling

  After random selection of first case, draw every nth case from population.

  Example: Every 5th patient or procedure.

- Stratified Random Sampling

  The population is divided into groups; each member of the group has an equal probability of being selected.

  Example: Divide population into particular disease categories and select from each group randomly.

Sampling designs

Types of non-probability sampling

- Convenience Sampling

  Any available group of subjects is used (lack of randomization)

  Example: Approaching smokers noted on the street to complete a survey on smoking.

- Snowball Sampling

  Subtype of Convenience Sampling

  Subjects suggest other subjects-

  Example: Asking cancer patients in a clinic to identify other cancer patients they know to be part of the sample.

Sampling designs

- Purposive or Judgement Sampling

  Useful for situations where you need to reach a targets sample quickly and where sampling for proportionality is not the primary concern.

  A particular group is subjectively selected based on criteria (also known as Cohort sampling)- Example: Using a group of nurses because researcher believes they represent a cross section of women.

- Quota Sampling

  A judgement is made about the most representative sample.

  Example: 25 charts per month or 5% or 30 - whichever is greater

Identify and Select Measures

Structure Measures

- Factors affecting context in which healthcare is delivered (physical facility, equipment , human resources, staff training, payment methods)
- Control how providers and patients in a healthcare system act
- Measures of average quality of care within a system Often easy to measure and observe
- Example: Are there enough staff to complete the project? Sufficient numbers of a certain type of equipment, ex. Sufficient numbers of medication scanners to reduce workarounds in a hospital or sufficient registration staff to process incoming people at a vaccination clinic.

Identify and Select Measures

Process Measures

- Sum of all actions making up healthcare (diagnosis, treatment, preventive care, patient education, actions taken by patients/families)
- How care is delivered (technical processes)
- Manner in which care is delivered (interpersonal processes)
- Measurement of process nearly equivalent to measure of quality
- Example: The percent of nursing home staff that attend COVID-19 vaccination education sessions. The number of audits completed at a managed care payer.

Identify and Select Measures

Outcome Measures

- Effects of healthcare on patients or populations (changes to health status, behavior, knowledge, patient satisfaction, health related quality of life). The result of the actions taken.
- Sometimes seen as most important indicators of quality because improving patient health status is the primary goal of healthcare
- Accurately measuring outcomes that can be attributed exclusively to healthcare is difficult
- Example: % of Nursing Home staff vaccinated for COVID-19.

Anexas Consultancy Services

Identify external data sources for comparison

Benchmarking

Involves asking the right questions

- What is the best practice?
- What are we doing?
- How are we doing it?
- How well are we doing it?
- What are the measurement results?
- Why are we looking for improvement?

Benchmarking

Potential data sources for benchmarking

- Government (CMS, CDC, state agencies)
- Large healthcare alliances or systems
- State peer review organizations and hospital associations
- For-profit database companies

Anexas Consultancy Services

Design scorecards and dashboards for different audiences

Information Dissemination

- Identify who needs to know information
- Determine what specific information they need
- Develop a system where the right people receive the right information at the right time in the right way

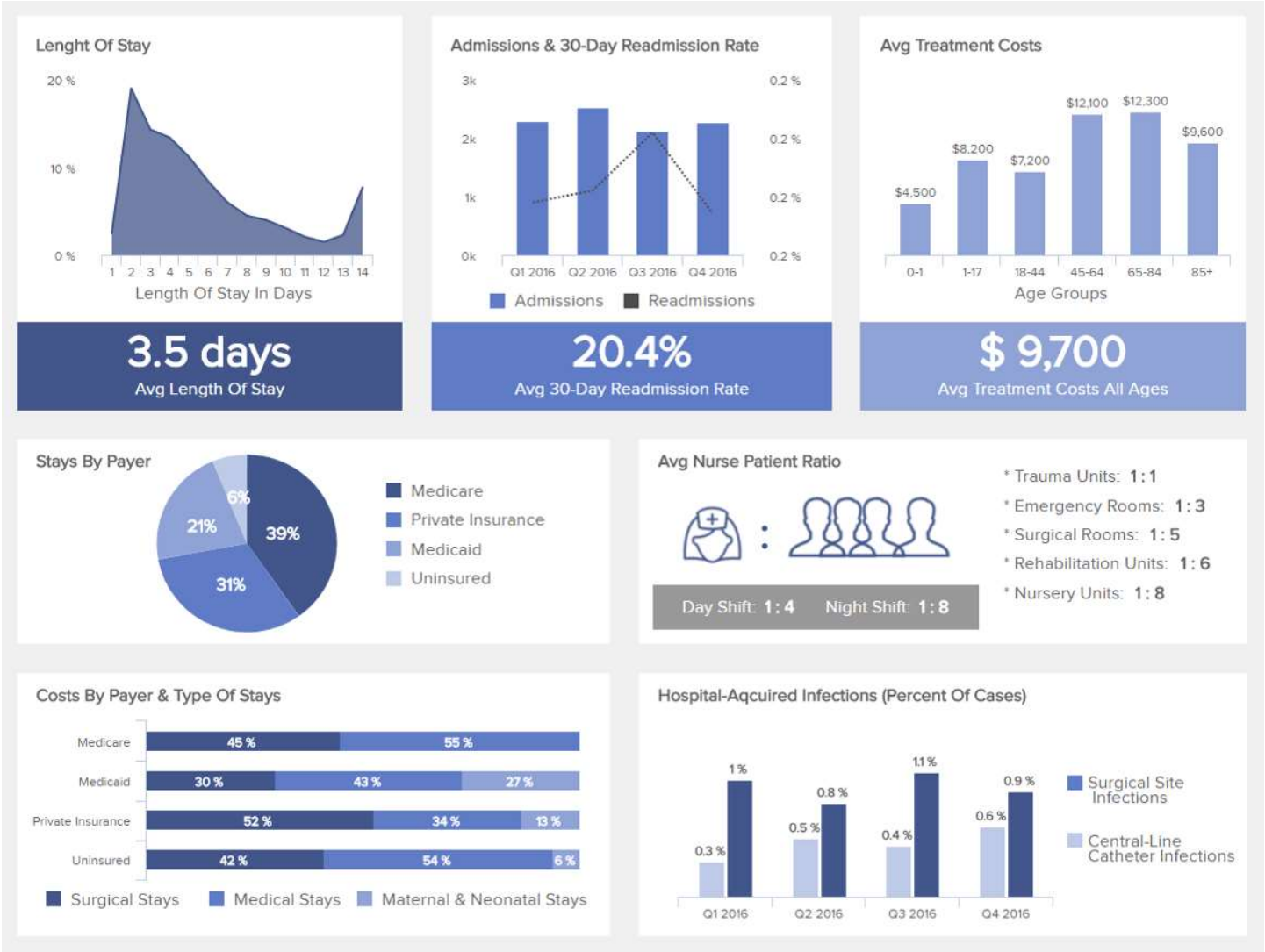
Dashboards and Balance Score Card

- Dashboards and balanced scorecards (BSC) are approaches to monitor the progress of performance against strategic goals.
- Dashboards and BSC can be used at organizational and/or departmental level since each department (unit) has its own goals and objectives.

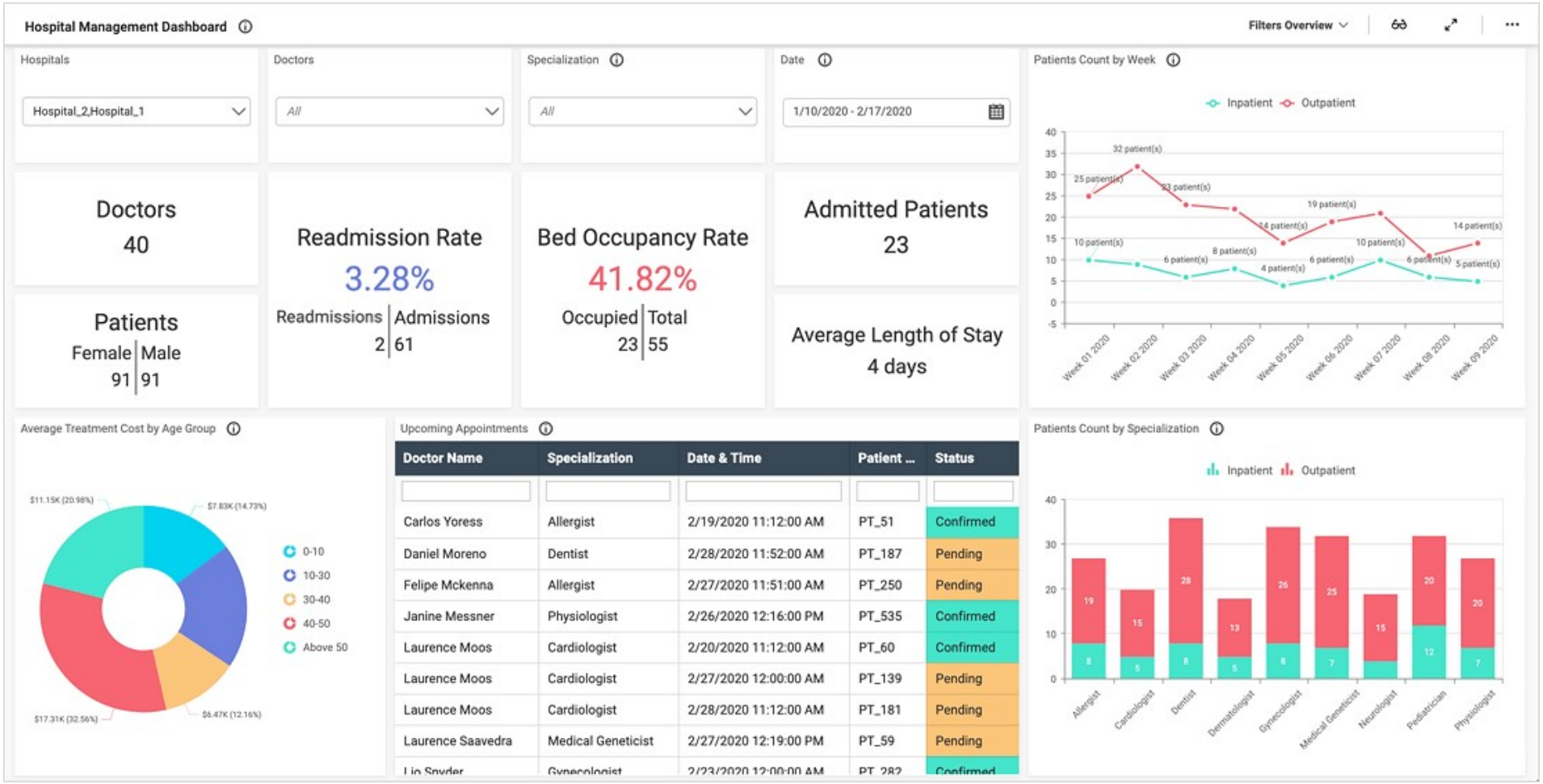
Dashboards

- Dashboards provide overview of an organization.
- Dashboards are an easy access computer program integrates a variety of performance /key indicators into one display, usually with graphs or charts.
- Dashboards provide a running picture of a departmental organizational status at any point of time.

Dashboard

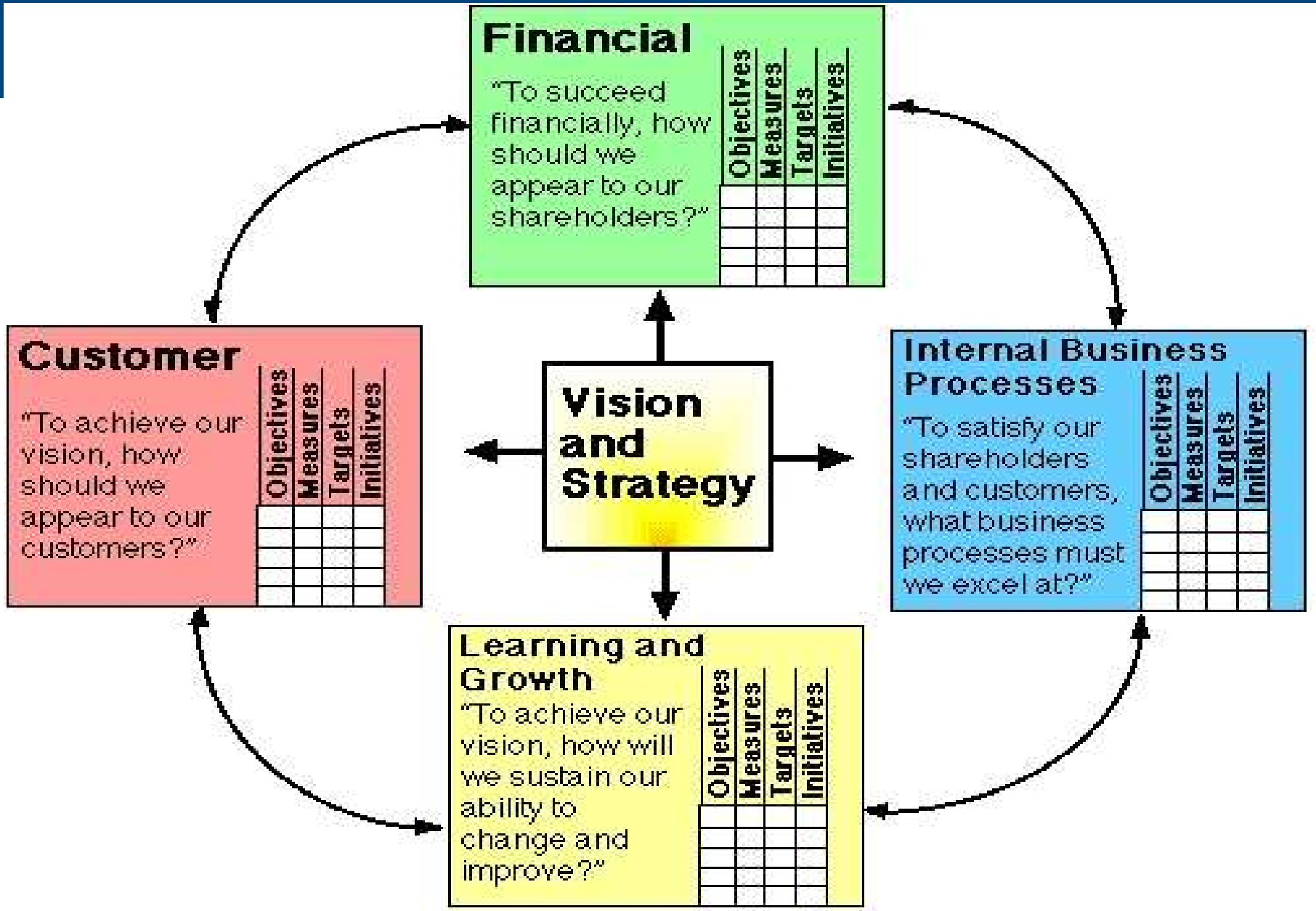

Lenght Of Stay
20 %
10 %
0 %
1 2 3 4 5 6 7 8 9 10 11 12 13 14
Length Of Stay In Days
3.5 days
Avg Length Of Stay
Admissions & 30-Day Readmission Rate
3k
2k
1k
0k
0.2 %
0.2 %
0.2 %
0.2 %
Q1 2016 Q2 2016 Q3 2016 Q4 2016
Admissions
Readmissions
20.4%
Avg 30-Day Readmission Rate
Avg Treatment Costs
$4,500
$8,200
$7,200
$12,100
$12,300
$9,600
0-1 1-17 18-44 45-64 65-84 85+
Age Groups
$ 9,700
Avg Treatment Costs All Ages
Stays By Payer
39%
31%
21%
6%
Medicare
Private Insurance
Medicaid
Uninsured
Avg Nurse Patient Ratio
Day Shift: 1 : 4
Night Shift: 1 : 8
* Trauma Units: 1 : 1
* Emergency Rooms: 1 : 3
* Surgical Rooms: 1 : 5
* Rehabilitation Units: 1 : 6
* Nursery Units: 1 : 8
Costs By Payer & Type Of Stays
Medicare 45 % 55 %
Medicaid 30 % 43 % 27 %
Private Insurance 52 % 34 % 13 %
Uninsured 42 % 54 % 6 %
Surgical Stays
Medical Stays
Maternal & Neonatal Stays
Hospital-Aqcuired Infections (Percent Of Cases)
0.3 %
1 %
0.5 %
0.8 %
0.4 %
1.1 %
0.6 %
0.9 %
Q1 2016 Q2 2016 Q3 2016 Q4 2016
Surgical Site Infections
Central-Line Catheter Infections

Dashboard


Hospital Management Dashboard
Filters Overview
Hospitals
Hospital_2,Hospital_1
Doctors
All
Specialization
All
Date
1/10/2020 - 2/17/2020
Patients Count by Week
Inpatient
Outpatient
Doctors
40
Readmission Rate
3.28%
Readmissions
2
Admissions
61
Bed Occupancy Rate
41.82%
Occupied
23
Total
55
Admitted Patients
23
Patients
Female
91
Male
91
Average Length of Stay
4 days
Average Treatment Cost by Age Group
$11.15K (20.98%)
$7.83K (14.73%)
$17.31K (32.56%)
$6.47K (12.16%)
0-10
10-30
30-40
40-50
Above 50
Upcoming Appointments
Doctor Name
Specialization
Date & Time
Patient ...
Status
Carlos Yoress
Allergist
2/19/2020 11:12:00 AM
PT_51
Confirmed
Daniel Moreno
Dentist
2/28/2020 11:52:00 AM
PT_187
Pending
Felipe Mckenna
Allergist
2/27/2020 11:51:00 AM
PT_250
Pending
Janine Messner
Physiologist
2/26/2020 12:16:00 PM
PT_535
Confirmed
Laurence Moos
Cardiologist
2/20/2020 11:12:00 AM
PT_60
Confirmed
Laurence Moos
Cardiologist
2/27/2020 12:00:00 AM
PT_139
Pending
Laurence Moos
Cardiologist
2/28/2020 11:12:00 AM
PT_181
Pending
Laurence Saavedra
Medical Geneticist
2/27/2020 12:19:00 PM
PT_59
Pending
Lio Snyder
Gynecologist
2/23/2020 12:00:00 AM
PT_282
Confirmed
Patients Count by Specialization
Inpatient
Outpatient
Allergist
Cardiologist
Dentist
Dermatologist
Gynecologist
Medical Geneticist
Neurologist
Pediatrician
Physiologist

Balance Score card

- BSC provide comprehensive view of organizational performance rather than depending on few choice indicators.
- BSC helps organizations to better link long-term strategy with short-term activities.
- BSC provide performance measures in relation to strategic plan (mission, vision, goals)
- BSC approach views the organization from four different perspectives:
  - *Financial (How do we look to providers of financial resources?)*
  - *Customer (How do our customers see us?)*
  - *Internal process (At what must we excel?)*
  - *Learning and growth (Can we continue improve and create value for customers?)*


Financial
"To succeed financially, how should we appear to our shareholders?"
Objectives
Measures
Targets
Initiatives
Customer
"To achieve our vision, how should we appear to our customers?"
Objectives
Measures
Targets
Initiatives
Vision and Strategy
Internal Business Processes
"To satisfy our shareholders and customers, what business processes must we excel at?"
Objectives
Measures
Targets
Initiatives
Learning and Growth
"To achieve our vision, how will we sustain our ability to change and improve?"
Objectives
Measures
Targets
Initiatives

Balance Score card

- In Healthcare, the most common adjustment to the traditional balanced scorecard is the extra emphasis on patient results and customer satisfaction

- If the scorecard is adequately balanced, it will reflect both the needs and priorities of the organization itself, and also those of the community and customers it services

Important Note:

The main difference between a dashboard and a scorecard is that a scorecard describes past performance, while a dashboard depicts performance in real time.

Balance Score card

QUALITY	SAFETY	PATIENT AND CONSUMER EXPERIENCE	POPULATION HEALTH
• COPD pathway utilization at 90% or above on all appropriate patient • 30-day mortality rates at or below expected • 90% or more leaders educated in PI tools by end of year	• Serious Safety Events under 20 events per year • Leapfrog Patient Safety Grade B or above • All Hospital Acquired Infection rates below expected as measured by the National Healthcare Safety Network (NHSN)	• CAHPS overall rating or above 60th percentile • 85% of hospice family member rate "exceed expectations' • Grievance days to closure less than 7 days/average	• 30-day readmission rate below expected • 90% of primary care patients have advance directives on file • Provide 8 or more community health education programs per year

Anexas Consultancy Services

Collect and validate quantitative and qualitative data

Data Types

There are two types of Data

Continuous (Variable)

- *It is continuous and quantitative in nature*
- *It can contain decimal point*
- *It has a unit*
- *Examples : Project completion time : 23.6 days*
- *Thumb Rule : Good sample size is 30-50 for studying variable data*

Attribute (Discrete)

- *It is qualitative in nature*
- *It has an attribute and various categories of an attribute, which can be*
- *described by discrete numbers:*
- *Example: Project status (Attribute) – Delayed, Not delayed (Categories)*
- *Gender (Attribute) – Male, Female (Categories)*
- *Each of the categories above can be described in terms of counts, for example 30males and 20 females. You cannot have 30.1 males!*
- *Thumb Rule : 200-1100 is a good sample size for attribute data*

Anexas Consultancy Services

Use data management systems for organization, analysis, and reporting of data

Data Dictionary

- A collection of names, definitions and attributes of the data elements being collected
- Helps avoid data inconsistencies between collectors
- Assists with the consistency in the collection as well as use of the data for the project team

Anexas Consultancy Services

Use statistics to describe data and examine relationships (e.g., measures of central tendency, standard deviation, correlation, regression, t-test)

What is Statistics?

Definition of Statistics

- Statistics is the science of collecting, organizing, analysing and interpreting data in order to make a decision.

Branches of Statistics

- The study of statistics has two major branches — descriptive(exploratory) statistics and inferential statistics.

- Descriptive statistics is the branch of statistics that involves the organization, summarization, and display of data.
- Inferential statistics is the branch of statistics that involves using a sample to draw conclusions about population. A basic tool in the study of inferential statistics is probability.

Descriptive Statistics

- Measures of central tendency or location (Mean, Median, Mode)
- Measures of Dispersion/Variability (Range, Standard Deviation, Interquartile Range, Variance)
- Measures of Position (Quartiles, Percentiles)
- Measures for **Categorical Data (Proportion, Percent, Rate, Ratio)**

Measures of Central Tendency

Mean :

- The most commonly used measure of location is the arithmetic mean.
- It is obtained by adding all the individual values in a population or sample together and dividing by the number in the group
- It measures the middle of the distribution.
- Properties:
  - ***Uniqueness:*** *For a given set of data there is one and only one mean.*
  - ***Simplicity:*** *Easy understood and computed.*
- The mean is affected by every value as it uses all values in the given set of data. Extreme values have an influence on the mean and in some cases, can so distort it that it becomes undesirable as a measure of location.

Measures of Central Tendency

Median

- The median is the middle number after they have been sorted
- It is the value which divides the set into two equal parts.
- Properties:
- **Uniqueness:** As is true with the mean, there is only one median for a given set of data.
- **Simplicity:** The median is easy to calculate.
- It is not drastically affected by extreme values as in case of the mean.
- Since it uses the middle value of the data set. The median is not a very reliable measure.

Measures of Central Tendency

Mode:

- The mode of a set of values is that value which occurs most frequently
- (E.g. 20,21,20,20,34,22,24,27 -> mode=20).
- If no fall values are different there is no mode.
- A set of value can have more than one mode.
- Properties:
  - *Simple: only the one most frequently occurring score.*
  - *Represents the highest bar in a histogram, or the highest point in a frequency polygon.*
  - *May appear in a distribution in places other than the center.*
  - *The only valid measure of central tendency for nominal data.*
  - *The least frequently used measure of central tendency as it does not lend itself to mathematical operations.*

Measures of Position

Quartiles:

- It is the value which divides the distribution into four parts subsets of equal size, each comprising 25% of the observations.
- To find the first, second, and third quartiles:
- Arrange the N data values into an array.
- First quartile, Q1 = data value at position (N + 1)/4
- Second quartile,Q2 = data value at position 1(N + 1)/4
- Third quartile, Q3 =data value at position 3(N + 1)/4

Percentile

- The values dividing the data into one hundred parts are called percentiles and are denoted by P1, P2, ... P99.

  It's the most common used among measures of position.

- E.g.: 70th percentile (P70) is the value such that 70% of the observations are less or equal to.

Measures of Dispersion (Variability)

- The measures of central tendency do not reveal the whole picture of the distribution of a data set. Two data sets with the same mean may have completely different spreads/variation.

- Example: MI patients weight: 53, 47, 60, 74, 66 & Pneumonia patients weight: 20, 85, 67, 36, 92

- The mean age of the two groups is 60 years, But, as we observe, the variation in the patient's weight in each of these two groups is very different. The weights of the pneumonia patients have a much larger variation than MI patients.

Spread

- It is a general statistical concept that describes the variability in the distribution of a quantitative variable.

- The measures of dispersion (variability) are measures of how different or spread the values of the variable of interest are from each other.

Range

- **Very simple;** it is just the difference between the highest and the lowest values.
- **Very rough;** it tells us the span over which the data are distributed.
- **The weakest;** It depends entirely on the extreme scores and do not tell us anything about the middle values.

Inter quartile Range

Quartile

Data is put in order and each quartile holds 25% of the total data components. Quartile gives some idea of the dispersion of data. There are three quartiles designated as Q1, Q2 and Q3

First quartile, Q1 = a value corresponding to 25%

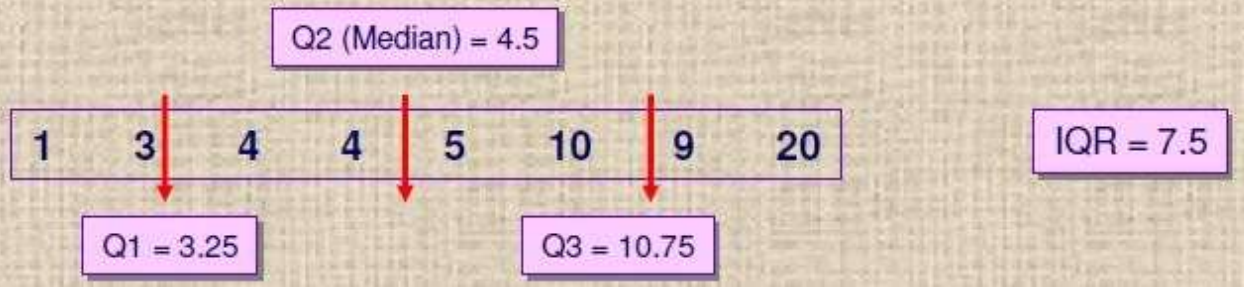
Second Quartile (median), Q2 = a value corresponding to 50%

Third Quartile, Q3 = a value corresponding to 75%

Inter quartile Range, IQR = Q3-Q1

Example - Calculate quartile and IQR for the data:

1, 10, 20, 4, 9, 5, 4, 3



Standard Deviation

- The most commonly used measure of spread is the standard deviation.
- The standard deviation is a measure of the spread of data about their mean.
- It tells us how close we are from the mean. A lower value of the standard deviation for a data set indicates that the values of that data set are spread over a relatively smaller range around the mean, and vice versa.

Standard Deviation

- Measure of variability - Average of deviations from the mean
- Standard: Average spread of scores around the mean
- Deviation: How much each score is scattered from the mean
- Most frequently used statistic for measuring degree of variability .σ is symbol for standard deviation called 'Sigma'
- A normal distribution is a standard bell curve
- Used with normally distributed interval or ratio data
- The greater the spread of the distribution, the greater the dispersion or variability from the mean (heterogeneous, more differences)
- The more values cluster around the mean, the smaller the variability or deviation (homogeneous, more similar)
- All scores are taken into consideration

Measures of Categorical data

- **Proportion:** is the number of observations with a given characteristic, e.g. in hydrocortisone group, the proportion of died patients = P= a/a+b = 50/200=0.25

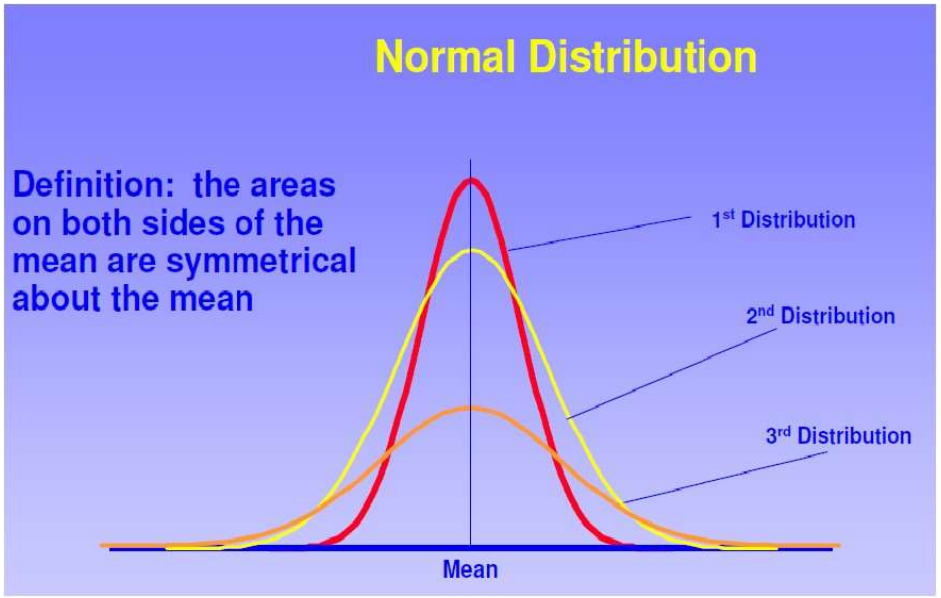
- **Percentage:** is a proportion multiplied by 100%, e.g. in hydrocortisone group, the percentage of died patients = (a/a+b) x 100 = (50/200) x 100 = 25%

- **Ratio:** is always defined as a part divided by another part. It is the number of observations in a given group with a given characteristic, e.g. in hydrocortisone group, the ratio of died patients to survived = a/b = 50/150 = o.33

- **Rate:** is similar to proportions except that a multiplier (e.g., 1000,10,000, or 100,000) called the base and usually computed over a specified period of time. Rate = (a/a+b) X Base. E.g. in hydrocortisone group, the rate of death per 10,000 patients per year = (50/200) x 10,000 = 2500 per10,000 patients per year.

Normal Distribution

- Normal distributions are a family of distributions that have the same general shape. They are symmetric with scores more concentrated in the middle than in the tails. Normal distributions are sometimes described as bell shaped.
- The normal distribution is easy for mathematical statistician to work with.



Inferential Statistics

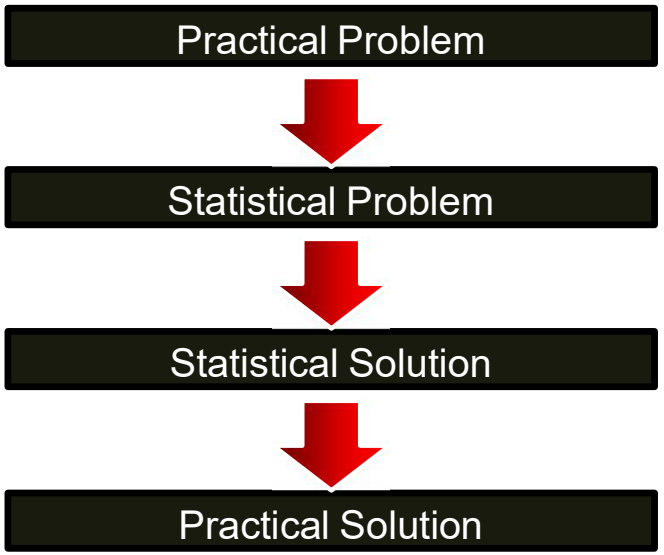
- **Hypothesis tests**
- **Confidence Level**
- **Confidence intervals**
- **Analysis of (ANOVA)**
- **Chi-Square Test**
- **Correlation and Regression Analysis**

Hypothesis tests

A procedure for testing a claim about a population parameter

We have two statistical hypotheses;

- ***Null hypothesis (H_0)* is hypothesis of no difference.**
- ***Alternative hypothesis (H_A),* is hypothesis of difference.**



Hypothesis tests

- **Student t-test (two-sample independent-group t-test):**

  It is the testing hypothesis that used for comparison between means of two independent groups.

- **Paired t-test: (two-sample dependent-group t-test):**
  - It is the testing hypothesis that used for comparison between means of paired groups.
  - Ways of pairing:
  - When subjects in one group are matched with similar subject in the second group.
  - When subjects serve as their own control by receiving both two different treatments.
  - When the same subjects are measured twice, in “before and after" studies.

Hypothesis tests

- ***Analysis of (ANOVA)***

  If we want to compare more than two groups;

- ***Chi-Square Test***

  The chi-square is the most commonly used methods for comparing proportions or percentages (categorical data, i.e. nominal) because it can be used to compare two or more independent proportions.

Correlation

- Correlation analysis and regression analysis are two procedures used to analyze association involving quantitative data, Although they share many similarities, correlation and regression have different objectives;
- Correlation analysis (Pearson's Correlation) is performed when the goal is to measure the strength or magnitude of the relationship between two numerical variables. Usually displayed graphically in a ***scatter plot.***
- Regression techniques are used to predict one quantitative variable from another quantitative variable.

Correlation Analysis

- Correlation: is a measure of linear association (linear, not causal). It needs two (Bivariate) continuous variables.
- Correlation coefficient (r) = - 1 to +1 (the number indicate the strength and the sign indicate the direction). Some time to prevent using (- or+), we may use ***coefficient of determination*** (r^2), then $0 \leq r^2 \leq 1$

Regression

- It is used when the aim of the study is to predict the value of one characteristic from knowledge of another. The goal of regression analysis is to derive a linear equation that best fits a set of data. This equation can be used to predict values of the dependent variable Y for given values of the independent variable X.

- **The straight line equation is: Y= a + bX**

- The higher the correlation between variables, the more accurate the degree of prediction. If there were no correlation between variables (r = 0), knowing the score of one would not help to estimate the score of the other.

Confidence Level

Is a percentage it means how confident you are

- Approximately 68% of your subject will fall within one standard deviation of the mean.
- Approximately 95% of your subjects will fall within two standard deviations of the mean.
- Over 99% of your subjects will fall within three standard deviations of the mean.

Confidence interval

- Confidence interval is a probability that the population parameter falls somewhere within the interval.
- It is calculated based on the mean, standard deviation and the sample size

Anexas Consultancy Services

Use data visualization and display techniques

Run Chart

Run chart is a chart that graphs data over time in relation to a central point. Such charts are superior to narrative since it shows a trend over time

Benefits of Run chart are;

- Identify trends and patterns in the process, like line graph.
- Identify how far we are from a central point, such as the mean or the median.
- Understand how a process is working and identify areas in need for improvement.
- Clarify if performance is static or changing (changing for better or worse?)
- Wait no interpret the results until at least 10 (or better, 20) data points have been plotted.

Run Chart

Four simple Rules to identify what Statisticians Patterns on Run Chart

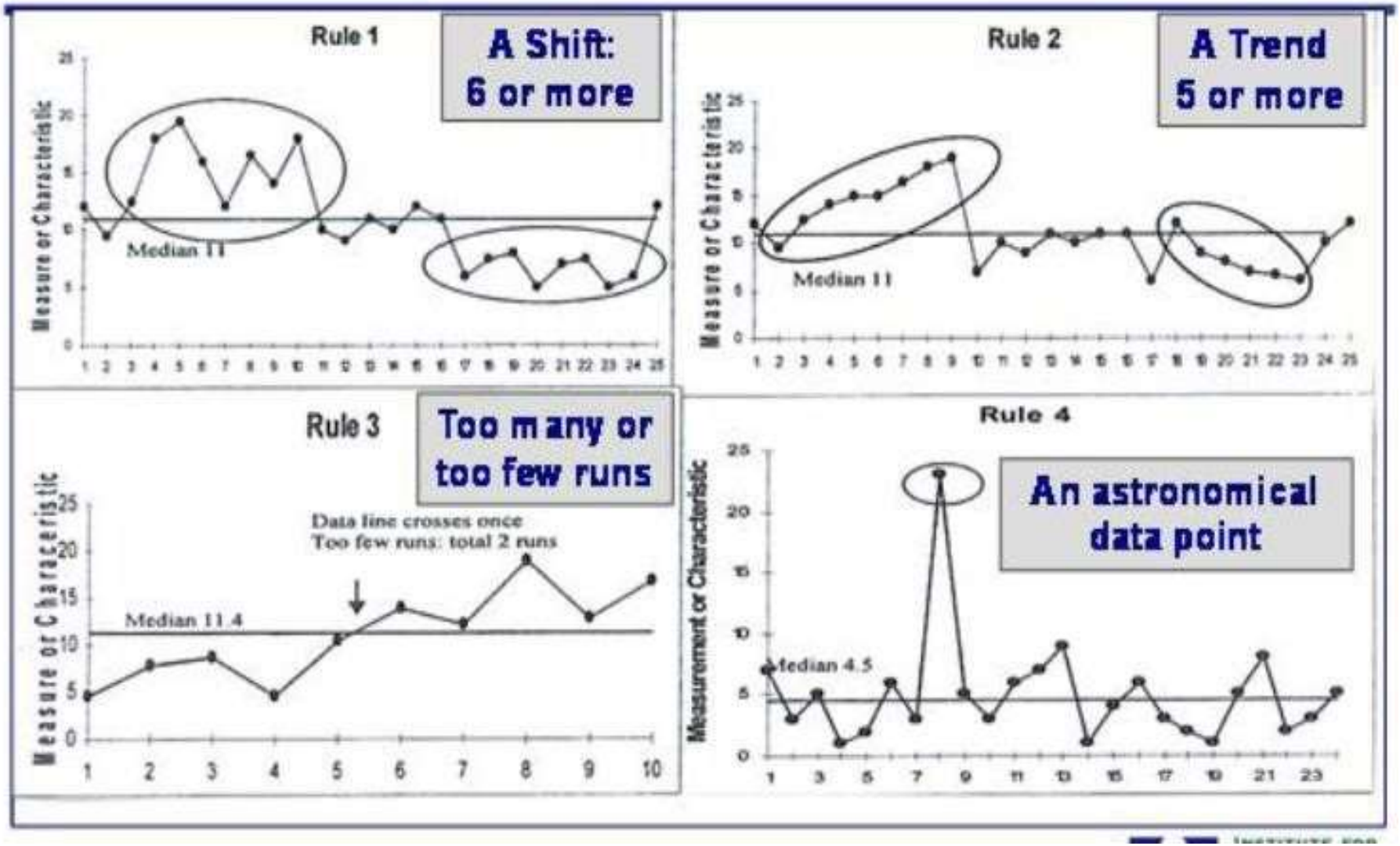
Rule 1: A shift in the process is indicated by six or more consecutive points above or below the median.

Rule 2: A trend is indicated by five or more consecutive points all increasing or decreasing.

Rule 3: Too many or too few runs indicate a nonrandom pattern. (Note: This one is complicated—we'll explain it in detail in another course.)

Rule 4: An "astronomical" data point is a pretty good signal of a nonrandom pattern.

How run chart determining area for improvement ?

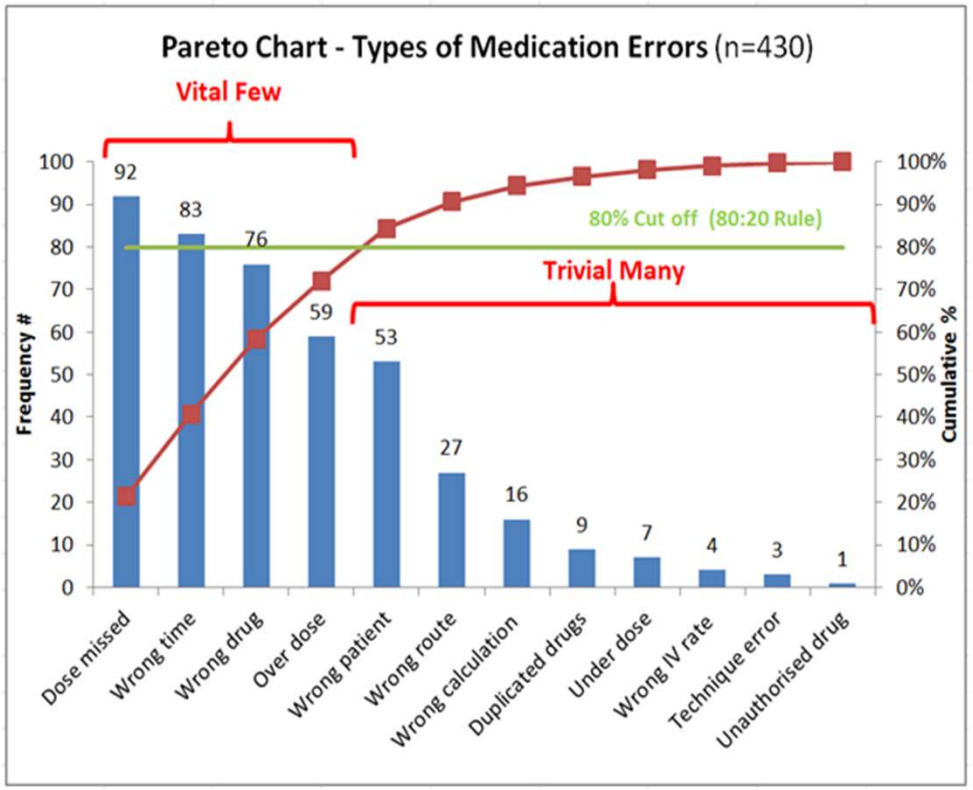


Pareto chart

- Pareto Chart is a type of chart which contains both bars and a line graph.
- It displays the values in descending order as bars & the cumulative totals of each category *as a line graph*.
- The left vertical axis is the frequency of occurrence and the right vertical axis is the cumulative percentage of the total number of occurrences. Because the reasons (Bars) are always in decreasing order, the cumulative function is a concave function.
- This ordering helps identify the **"vital few"** (the factors that warrant the most attention) from the **"useful many"** (factors that, while useful to know about, have a relatively smaller effect).

Pareto chart

- Using a Pareto diagram helps a team highlight the most important among large set of factors. In quality control, if often represents the most common sources of defects.
- This tool is related to Pareto (80/20 rule) which states that 80% of the problems come from 20% of the causes.



Pie Chart

What is a Pie Chart?

- A Pie Chart graphically displays categorical data as portions of a whole.
- Useful for understanding all the responses on a measure
- Usually expressed as percentages
- Slices of the pie represent percentage in each group

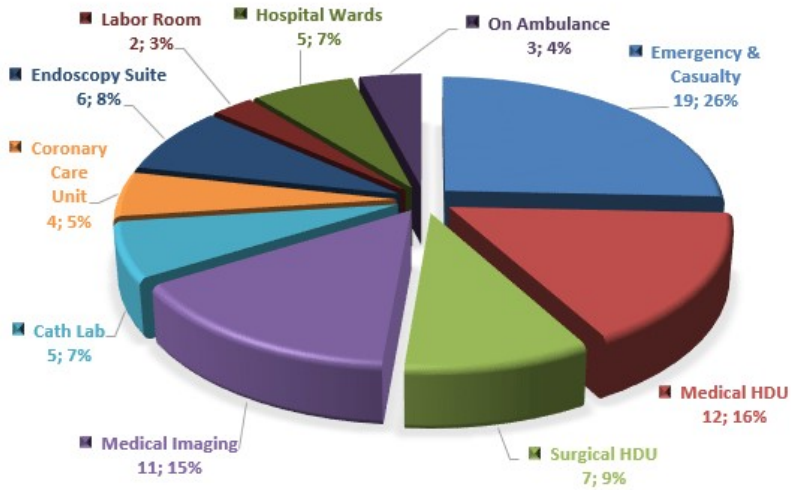
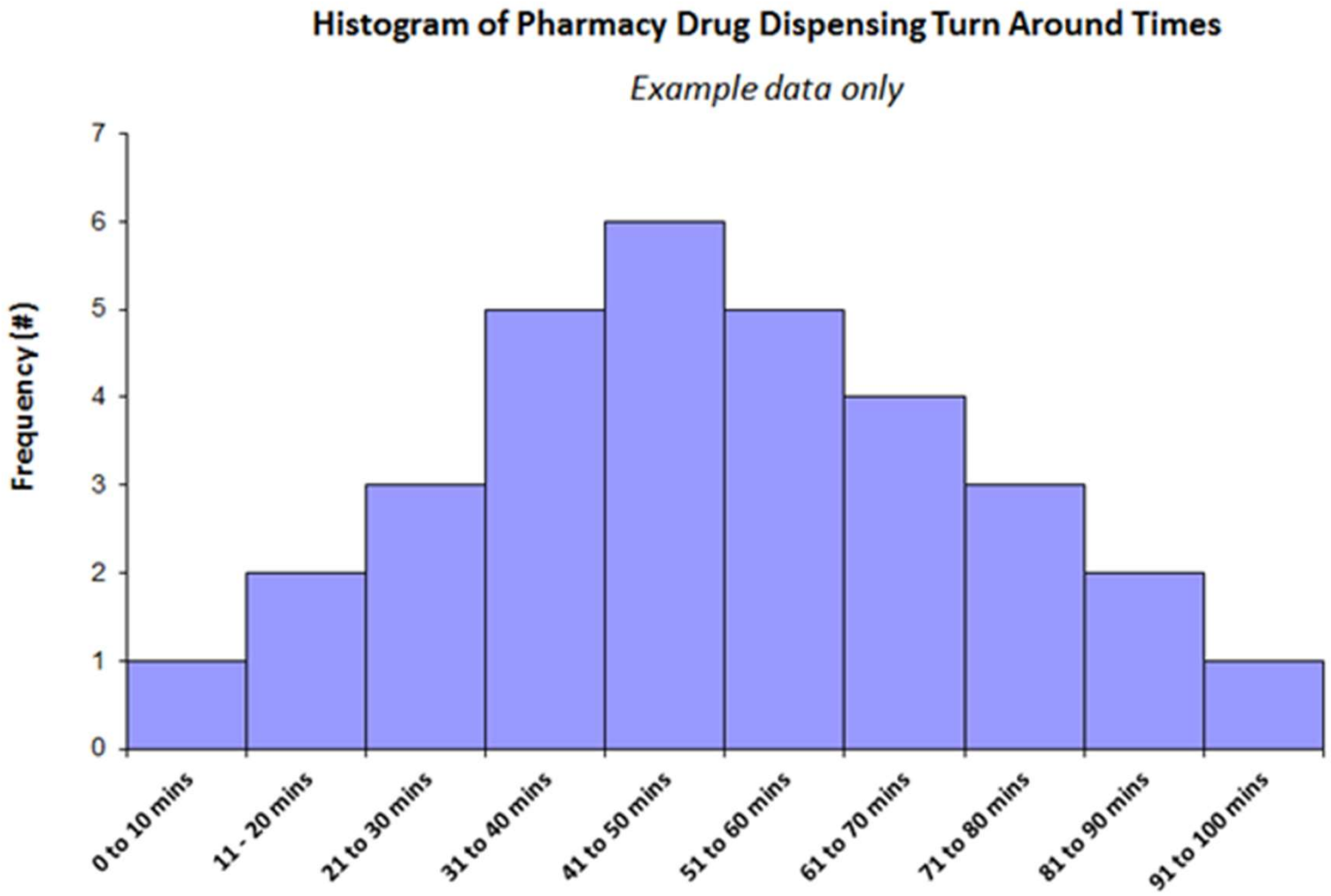


Figure-1: Pie-chart showing the number & percentages of patients attended at different sites (n=74). Note: Data of Critical Care Centre is not included. Percentages are shown in round figures.

Histogram

- A bar chart for one variable
- Used to visualize central location, shape and spread of the data
- Each bar equal, distinct - no overlap
- Most often used with time, money, throughput or a scaled measurement such as dollars, weight, height, age
- Will not show stability of a process To be most useful must have 50 data points or more
- The histogram illustrates the shape (or distribution) of the data by indicating how often different values appear
- High bar indicates mean , median , mode No space between bars

Histogram Example



Scatter Diagram

When to Use a Scatter Diagram ?

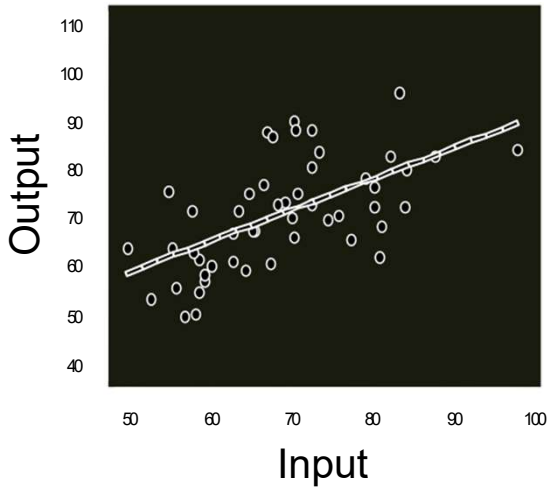
- When you want to test whether the performance of one factor (or variable) is related to the performance of another factor
- One factor is the independent (or driver) variable
- Second factor is the dependent (or follower) variable
- The dependent variable is the one that can be influenced
- Must have meaningful paired samples
- The outcome variable and the related suspect variable must be for the same patient

Scatter plot

- The Scatter Plot or Scatter Diagram is an XY plot.
- Points are plotted at the intersection of two data points.
- Data points are used for finding possible cause and effect relationships.

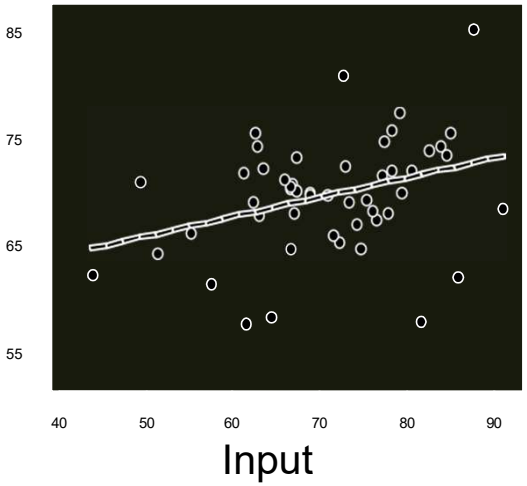
Positive relation

Moderate positive correlation



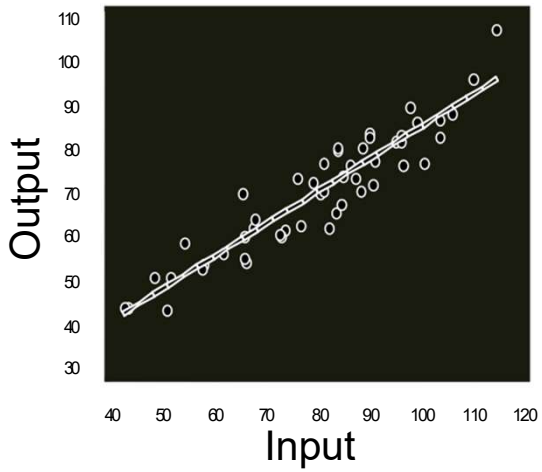
Y=25.7595+0.645418X
R Squared=0.369

Weak positive correlation



Y=56.6537+0.181987X
R Squared=0.115

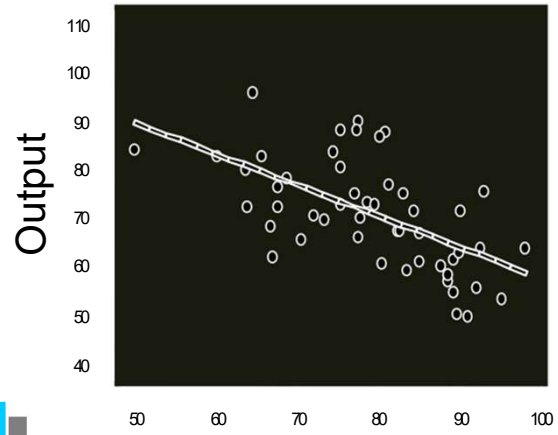
Strong positive correlation



Y=9.77271+0.745022X
R Squared=0.876

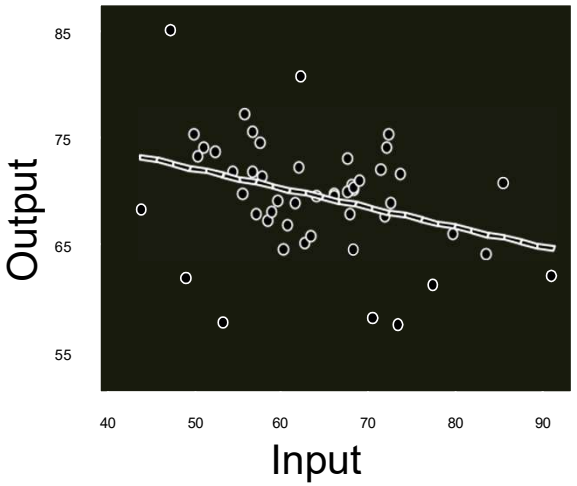
Negative relation

Moderate negative correlation



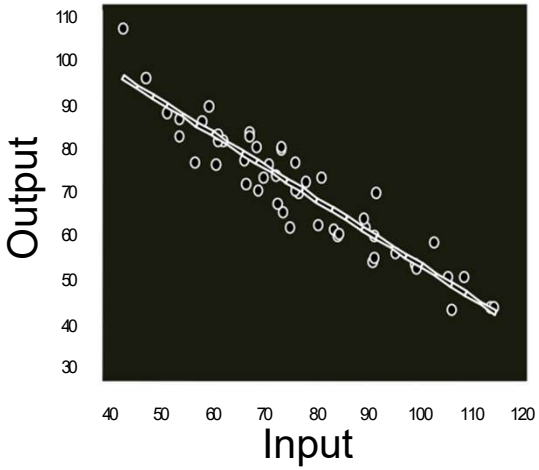
Y=90.3013-0.645418X
R Squared=0.369

Weak negative correlation



Y=74.8524-0.181987X
R Squared=0.115

Strong negative correlation



Y=99.1754-0.745022X
R Squared=0.876

Anexas Consultancy Services

Use statistical process control techniques and tools

Control Chart

- **Control Chart** “Shewhart Chart or Process Behavior Charts"
- Control chart is like run chart but it includes statistically determined control limits on each side of the center line (mean). These control limits are equal to +3SD and -3SD.
- Control chart is designed to identify;
- 1. Trends and patterns.
- 2. The type of variation (common cause or special cause).
- 3. The process is statistically within the control limit or not.

Control Chart

- Control charts are line graphs that are similar to Run Charts.
- Control Charts use Mean, and upper and lower limits, to determine special or common cause.

Use Probability-based rules for:

- Determining the stability of a system.
- Monitoring performance.
- Finding evidence of improvement.
- Evaluating sustainability of changes.

Using a Shewhart Chart

- Assess stability of a process
- Identify actions needed to have a positive impact on the data
- Find and evaluate causes of variation
- Control charts are superior for monitoring data
- Identifies if a change has led to an improvement
- See if improvements are "sticking"

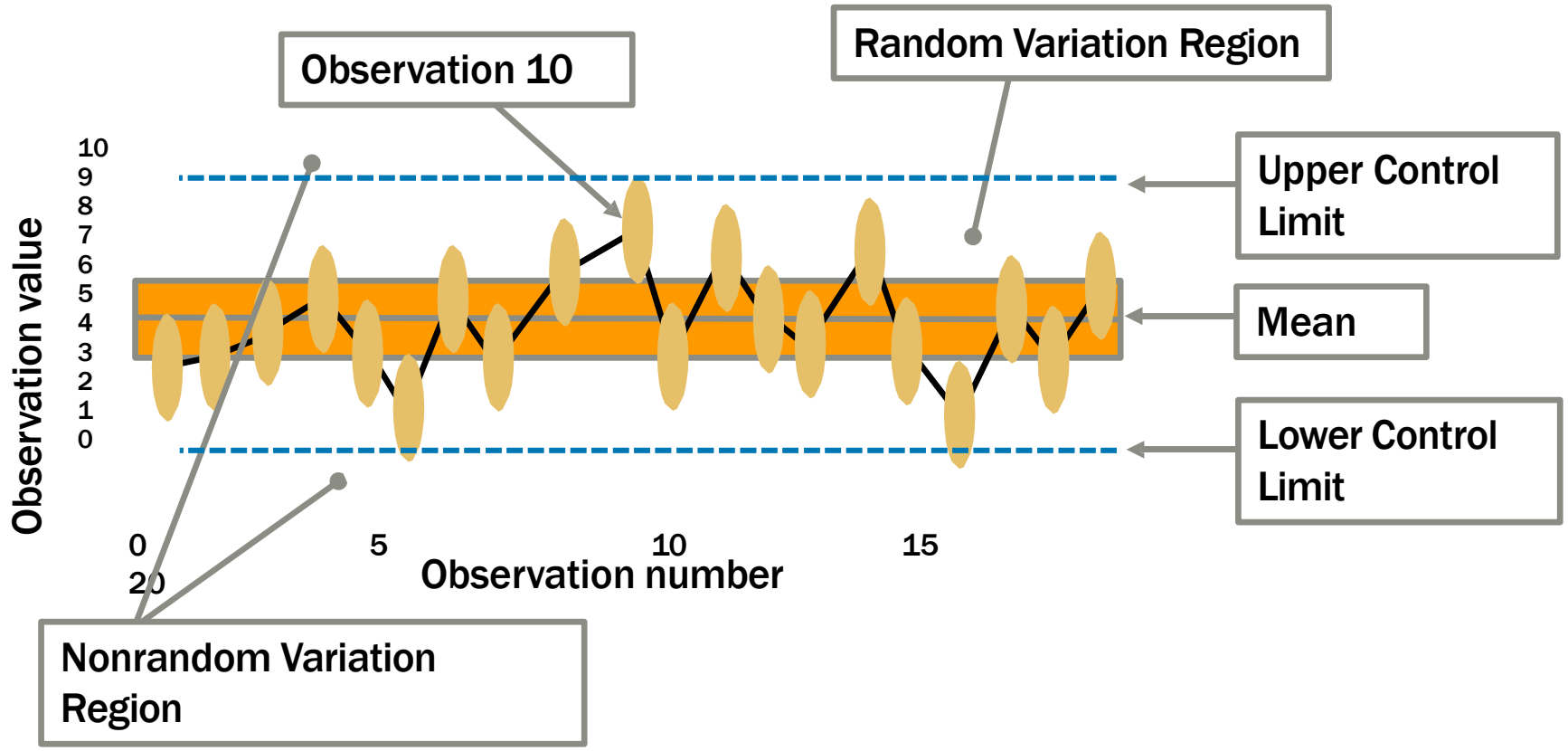
Control Chart

- Control charts rules to determine common vs. special cause variation:
- Common cause:- exhibited as points between the control limit and no particular pattern.
- Special cause:- exhibited as points are fall outside the control limits or when inside the control limits with one of the following pattern:

- Note:- the aim is to reduce common cause variation and eliminate special cause variation.

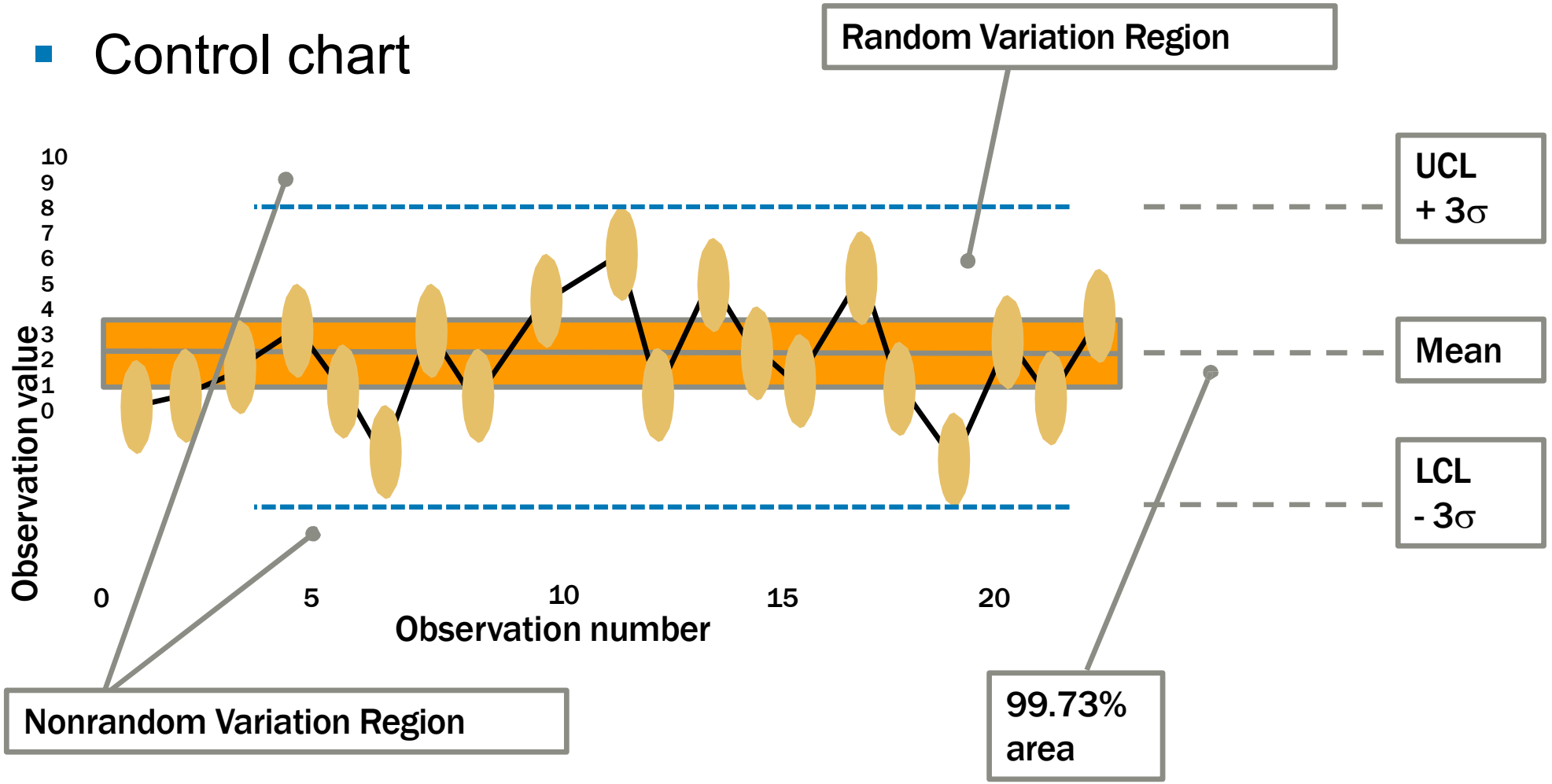
Components of Control chart

- Control Chart

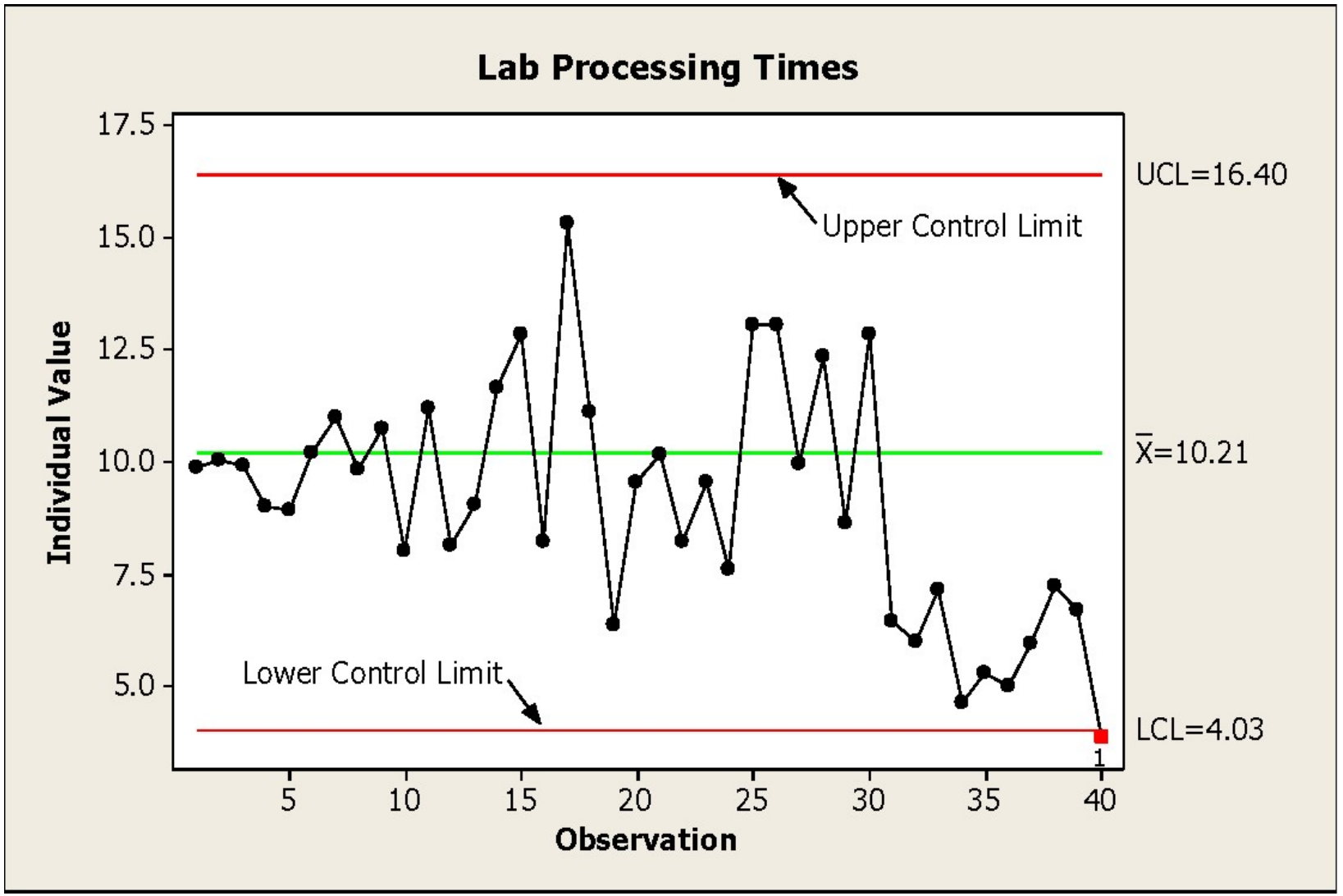


Statistics of Control chart

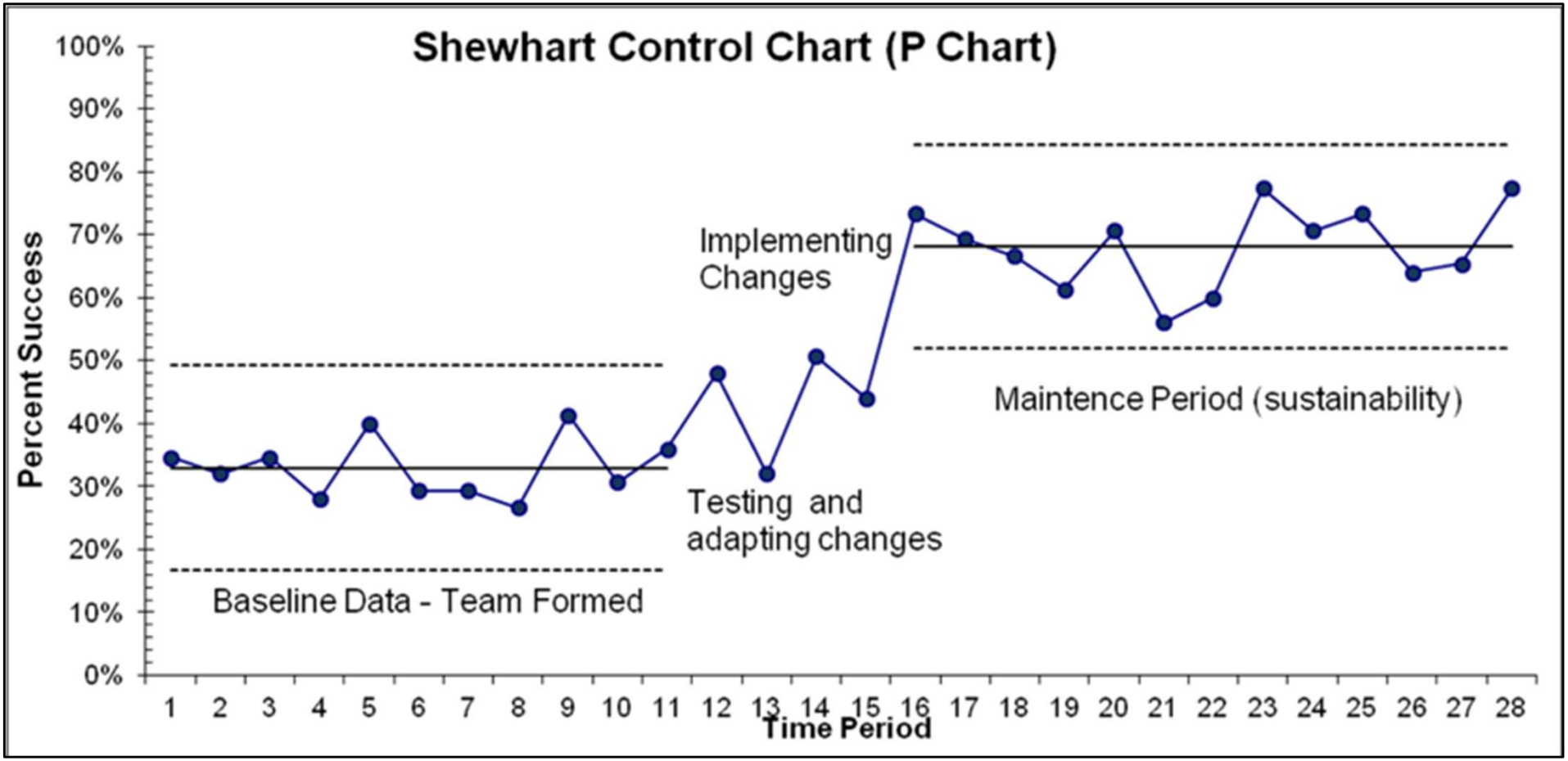
- Control chart



Control chart



Control Chart



Anexas Consultancy Services

Use measurement tools to evaluate process improvement

Measurement Tools : Reliability and Validity

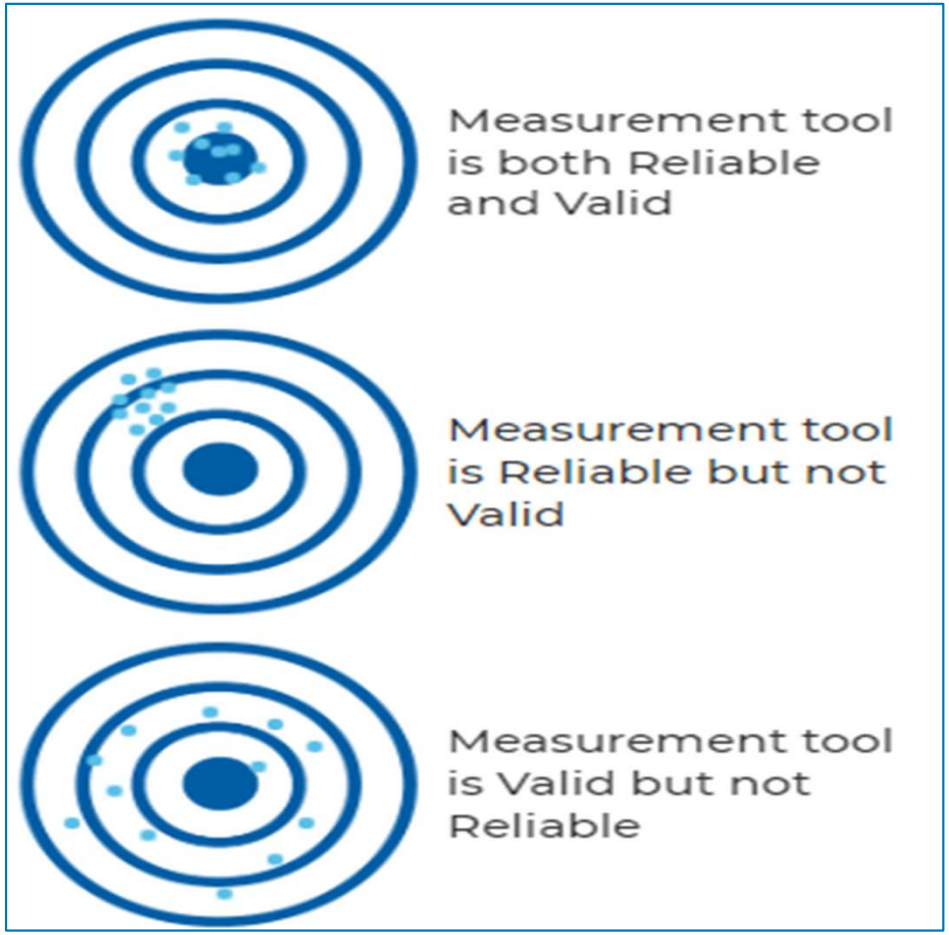
Reliability

- Extent to which an instrument yields the same result on repeated trials.
- Tools used include Test-retest, Split half, Reliability by equivalence

Validity

- Degree to which instrument measures what it is intended to measure
- More difficult to establish than reliability.
- An instrument that is not reliable cannot be valid. An instrument may be reliable but does not have to have validity.

Reliability and Validity



Types of Reliability

- TEST-RETEST
- SPLIT-HALF
- RELIABILITY BY EQUIVALENCE
- INTER-RATER RELIABILITY

TEST-RETEST

- Determines the stability of an instrument by administering a test to a sample of people on two occasions and comparing the scores.
- Example: Test-retest - If you weigh yourself multiple times and get the same results, then the scale is reliable

Types of Reliability

SPLIT-HALF

- Split half assesses internal consistency by correlating scores on half of the test with the scores on the other half.
- A measure of consistency where a test is split into two and the scores for each half of the test are compared with one another
- If the results are similar, demonstrates reliability by equivalence
- Example: The CPHQ Comprehensive Pre-test is split into two halves and taken by the same group of people. This demonstrates consistency and reliability of the tests.
- Numerical result compares the two results for the Reliability Coefficient.

Types of Reliability

RELIABILITY BY EQUIVALENCE

- Numerical index of test's reliability
- The closer the coefficient is to 1.0, the more reliable the tool.
- Generally a reliability coefficient of ≥ 0.70 are acceptable, although ≥ 0.80 is preferred
- Calculation of the reliability coefficient is not within the scope of the CPHQ exam.

Types of Reliability

INTER-RATER RELIABILITY

- Degree to which two raters , operating independently, will assign same ratings
- Example: A team of wound care specialists rate patient wound healing using a rating scale. The team compares the results for the same patient. When there is a strong correlation between the provider's results, there is a high interrater reliability.

Types of Validity

Content Validity

- Degree to which instrument adequately represents universe of content.
- Includes judgments by experts about the degree to which a test appears to measure what it is supposed to measure.

EXAMPLE

- The content validity of the CPHQ exam: A test has content validity if it measures knowledge of the content domain of which it was designed to measure knowledge. Another way of saying this is that content validity concerns the adequacy with which the test items adequately and representatively sample the content area to be measured.

Types of Validity

Criterion Validity

- Extent that score on instrument can be related to a criterion (behavior instrument is supposed to predict)

EXAMPLE

- Patient acuity index:
- This tool uses criterion about the patient population on the unit to assign the appropriate number and skill set of staff to a patient care unit.

Summary

- In this module you used different Information System types such as Administrative Support Systems, Clinical Information Systems, and Decision Support Systems to decide the best candidate for the Quality Manager position.
- With the help of your new Quality Manager, you used different sampling designs to determine the correct sample group to use for your data collection and study.
- Your new Quality Manager and the other members of your team identified and selected the best measures, Structure Measure, Process Measure, and Outcome measure, to use to decrease the HgbA1C in your patients with diabetes.
- You and your team selected the best organization to use as a benchmark for chemotherapy treatments for cancer patients at your organization.
- You and your team used different types of data and statistical power to study and display data about patients with hypertension.
- Your new Quality Manager guided the team in working with various data display types such as Run Charts, Pareto Charts, Histograms, Scatter Diagrams, Pie Charts, and Control Charts to display and interpret data.