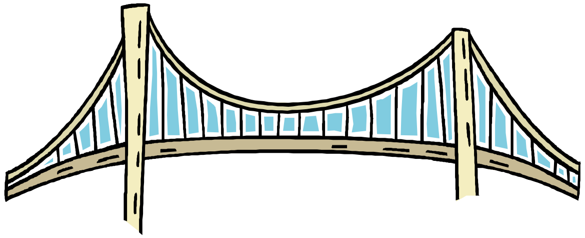
Continuous Improvement Toolkit

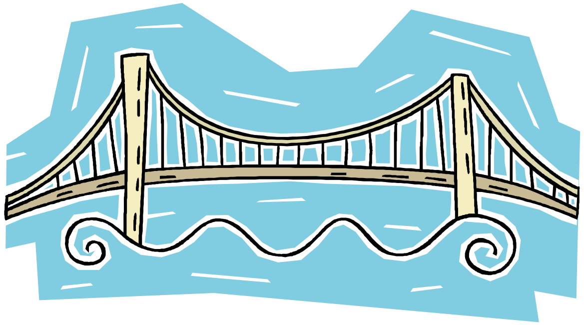
Probability Distributions



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

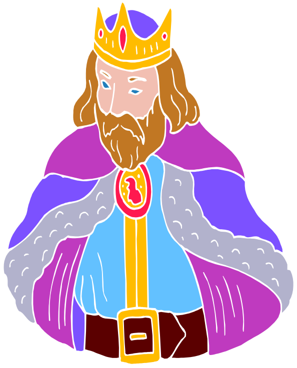
- Probability Distributions

- Probability distributions provide the bridge between descriptive and inferential statistics.
- They are mathematical models that we use to model real-life data distribution.
- Once we have found the appropriate model, we can use it for **prediction purposes**.

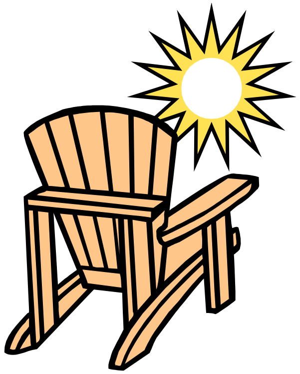


- Probability Distributions

- Probability runs on a scale of 0 to 1.
- A probability of 0 means ‘it will definitely not happen’.
  - **Example:** what is the probability that you will be the king of England?

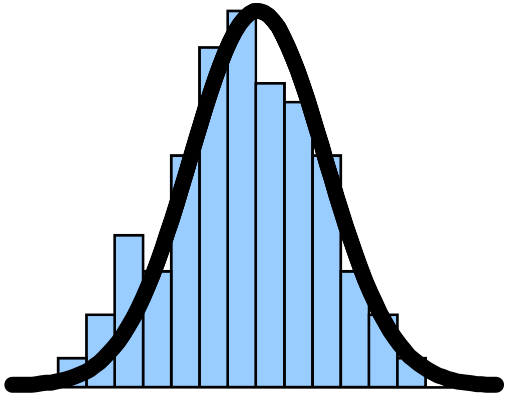


- A probability of 1 means ‘it is certain to happen’.
  - **Example:** what is the probability that the sun will rise tomorrow?



- Probability Distributions

- When dealing with process improvement, we generally have only **samples** to work with.
- A sampling distribution is the probability distribution of a given statistic based on a random sample.
- To draw conclusions from sample data, we compare values obtained from our sample with the theoretical values obtained from the probability distribution.
- We need to be sufficiently confident before we take any decision.
- This “Confidence level” is often set at 95% or 99%.



- Probability Distributions

Probability Distributions Types - Discrete Distributions:

- Deals with data that can take specific values.
- Count and Attribute data worlds.
- Two common discrete probability distributions: Binomial & Poisson.
- **Examples:**
  - The number of defective items in a sample.
  - The number of defects found in a single product.

Binomial

Poisson

Probability Distributions Types - Continuous Distributions:

- Relate to data which can take any value.
- There are a number of different distribution models and shapes.
- The commonest is the Normal distribution.
- The Exponential and Weibull distributions are widely used in the field of reliability engineering.

F

Uniform

Weibull

Student's T

Normal

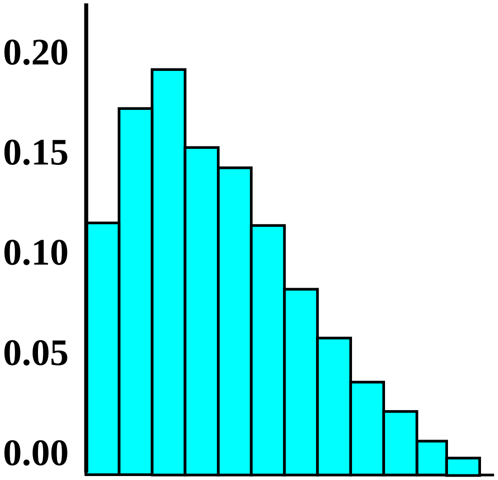
Exponential

- Probability Distributions

The Binomial Distribution:

- Used for data which can only take on of two values: pass/fail, yes/no, etc.
- **Example:**
  - Taking 15 samples from a large batch which is say 4% defective.

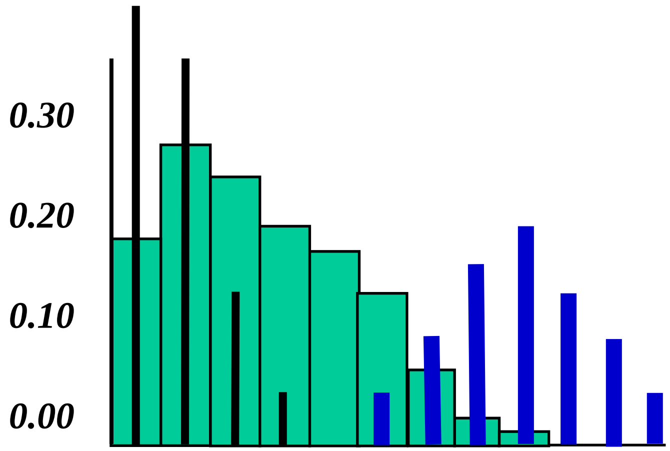
$$P(r) = \binom{n}{r} p^{r}(1-p)^{n-r}$$



- Probability Distributions

The Poisson Distribution:

- Used for data where there may be any number of defects in a given item.
- There must be a defined interval or area of opportunity (e.g. per sample).
- **Examples:**
  - The number of incoming calls per hour to a call center.
  - The number of failures per month for a specific equipment.
  - The number of accident per year in a factory.
- The exact shape depends solely on the value of λ.

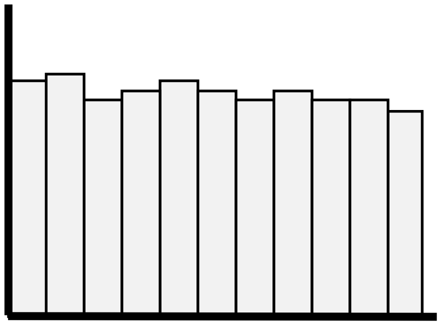


$$P(r) = \frac{\lambda^{r} e^{-x}}{r!}$$

- Probability Distributions

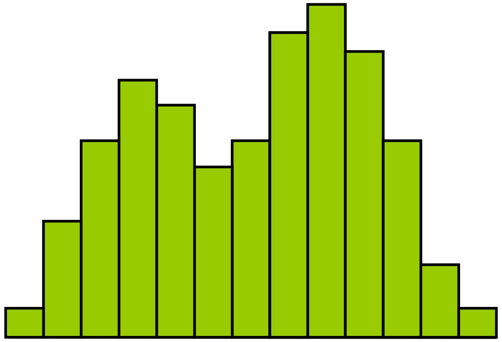
The Uniform Distribution:

- Describes variables that have a constant probability.
- Does not occur often in nature, but it is important as a reference distribution.
- An example of the discrete uniform distribution is throwing a dice.
- The possible values are 1, 2, 3, 4, 5, 6, and each time the dice is thrown the probability of a given score is 1/6.



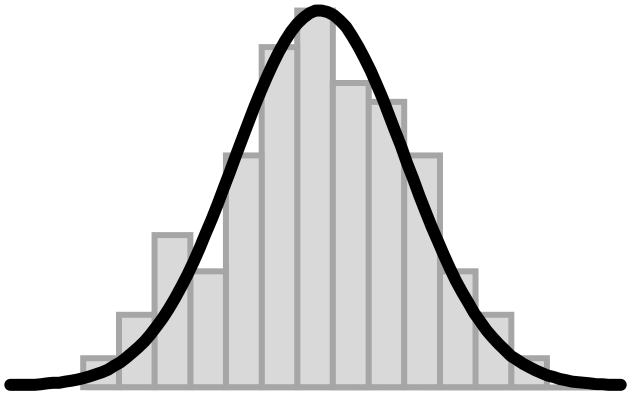
The Bimodal Distribution:

- Continuous probability distribution with two different modes.
- If observations are taken from different populations, multi-modality may result.
- **Example:**
  - Taking samples from two shifts.
  - The operators may setup the machine up differently.
- A **multimodal distribution** is a continuous probability distribution with two or more modes.



The Normal Distribution:

- A Normal distribution is a commonly occurring distribution.
- Natural phenomena, physical quantities and industrial processes follow approximately the Normal distribution.
- Often used in the natural and social sciences.
- If the normal distribution is applicable, it can be used to estimate future process performance.
- **Examples:**
  - Product thickness and weight.
  - Water purity and temperature.
  - Dimensions and test results.



- Probability Distributions

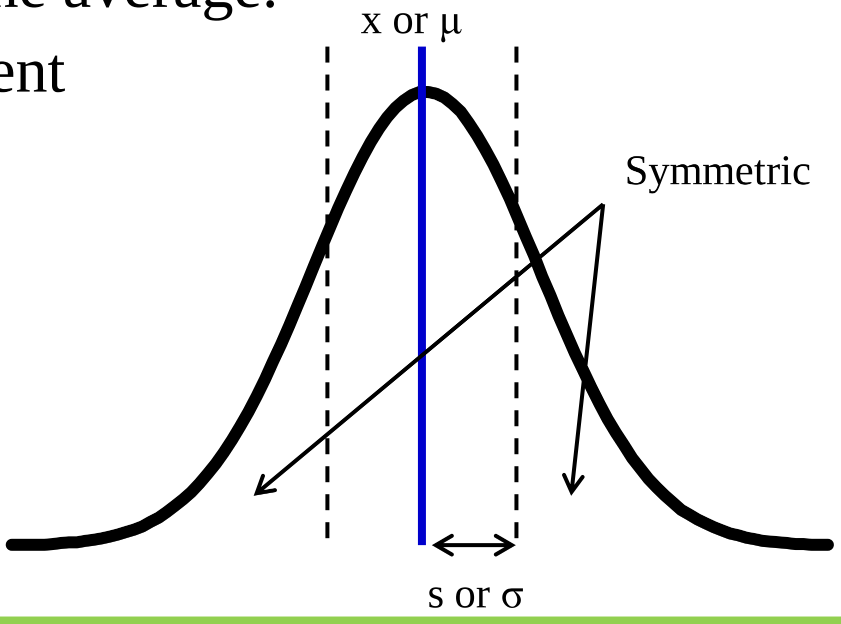
The Normal Distribution:

- It is symmetrical with most of the results in the middle and fewer toward the extremes.
- Fully defined by its mean and standard deviation.
- The peak of the curve represents the average.
- The spread of the curve is equivalent to six times the standard deviation of the process.

Symmetrical

Bell shaped

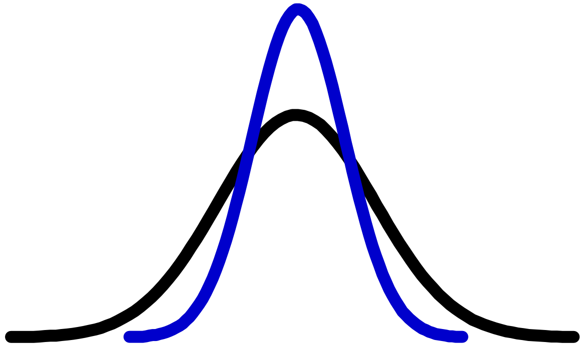
Centred on the mean



- Probability Distributions

The Normal Distribution:

- Uses a continuous curve to describe probabilities.
- The normal curve is based on the average and standard deviation of the histogram.
- A wider-flatter curve demonstrates more variation in the process.
- In theory, the normal distribution never ends but in practice, most results fall within +/– 3 Sigma.

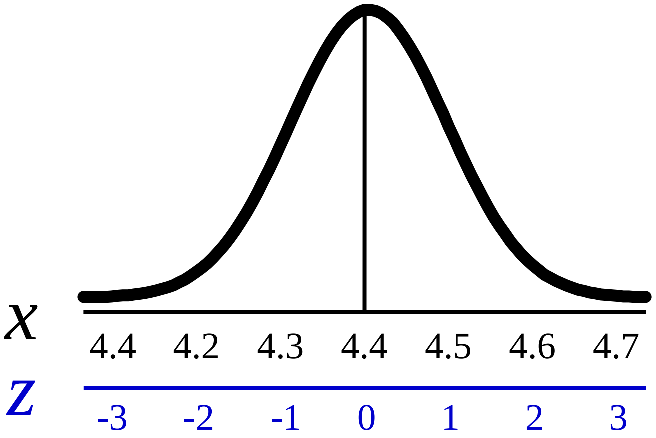


- Probability Distributions

The Standard Normal Distribution:

- Has a mean = zero, and a standard deviation = 1.
- Any normally distributed data can be converted to standard normal distribution.
- For any point 'x', the 'Z-value' is the number of standard deviations of that point from the mean.
- The area between two values gives the proportion of results that will occur between the two points.
- This calculation can give us a sense of the probability to make inferences about a population.

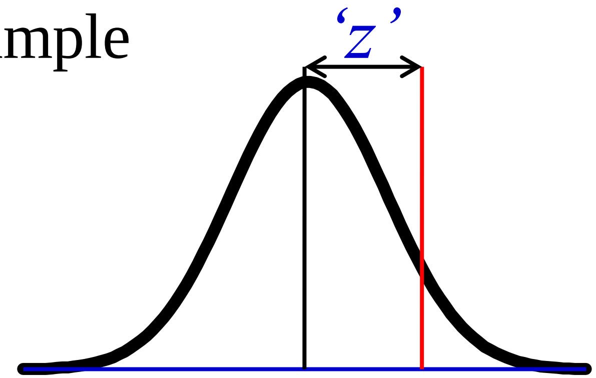
$$Z = \frac{x - \text{Average}}{\text{STDev}}$$



- Probability Distributions

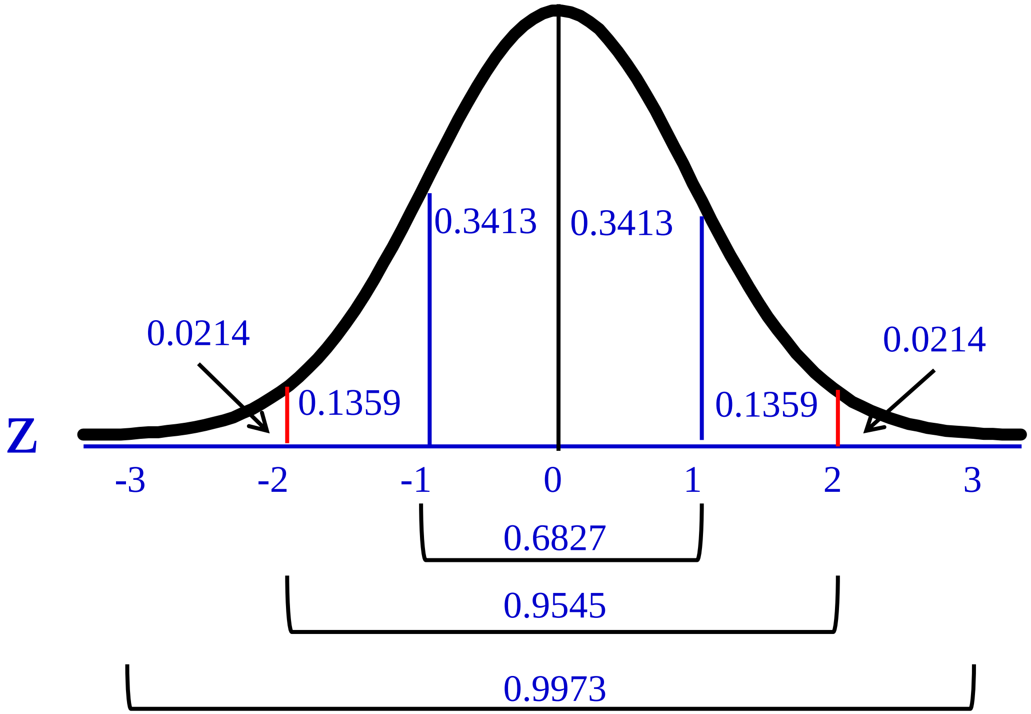
The Z-table:

- The complex shape of the normal curve has been converted into a mathematical table called the Z-Table.
- Used to find the probability that a statistic is observed below, above, or between values on the standard normal distribution.
- It is a common practice to convert a normal to a standard normal and then use the standard normal table to find probabilities.
- Using the Z-table, we can determine for example the proportion of outputs greater that the process average.



- Probability Distributions

- The Z-value measures the sample distance from the population mean, calculated in standard deviations.

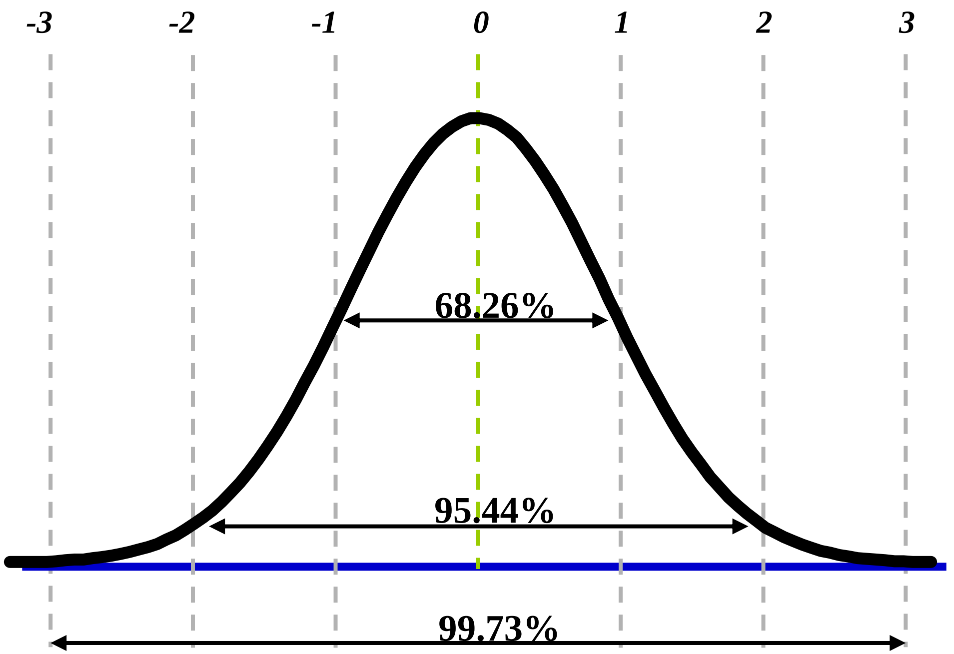


The total area under the normal curve equals to 1

- Probability Distributions

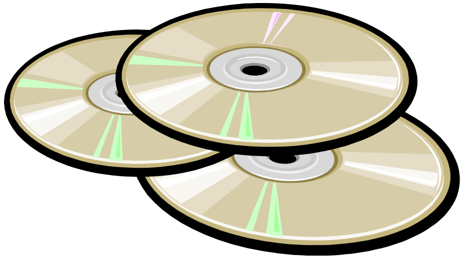
- If the normal distribution is applicable, it can be used to estimate future process performance.

For a Normally distributed output, 68% of the results will be within +/– Sigma of the average



- Probability Distributions

Example:

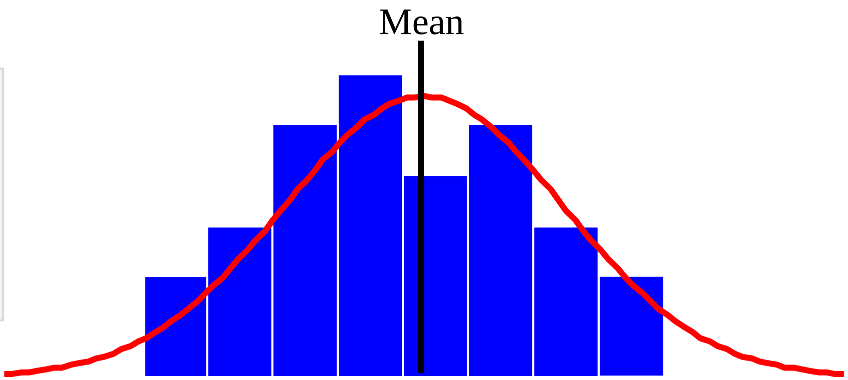


- A test run of **100** CDs has been made and the thickness results has the average of **1.35mm** and a standard deviation of **0.06**.
- None of the tested CDs were above the upper specification limit of **1.5mm**. However the Normal curve on the histogram still extends beyond.
- This suggests that over the long-term some CDs will be oversize.
- Given that the Normal distribution can be used, what is the probability that we will have defected CDs in the long term?
- The Z-table predicts the area under the curve to be 0.6%.
- This is different and a better prediction than the 0% predicted earlier.

- Probability Distributions

- **Several tools are available to help test data for normality:**
  - Histograms.
  - Normal probability plots.
  - Anderson-Darling normality test.

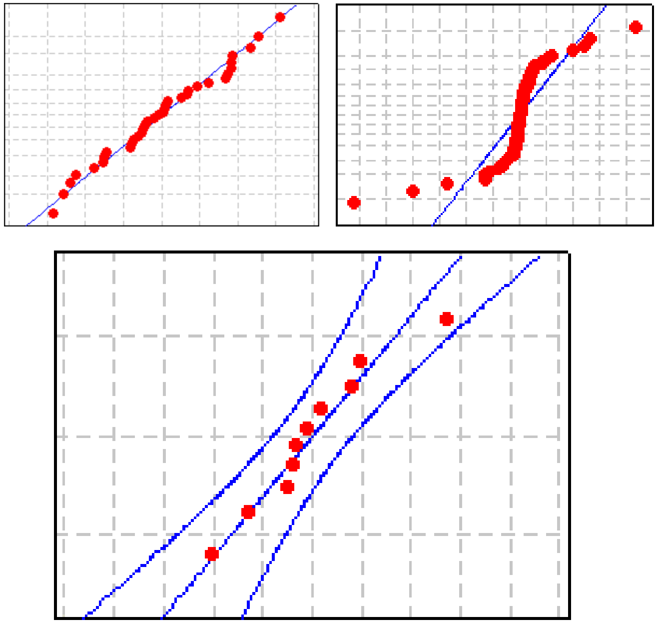
Histograms are efficient graphical methods for describing the distribution of data.



- Probability Distributions

Probability Plots:

- Provide a more decisive approach for deciding if a data set fits the normal distribution.
- Constructed in a way that the points will fall in a straight line if they fit the distribution question (e.g. Normal).
- A Normal distribution will form a straight line that falls between the 95% the CI limits.

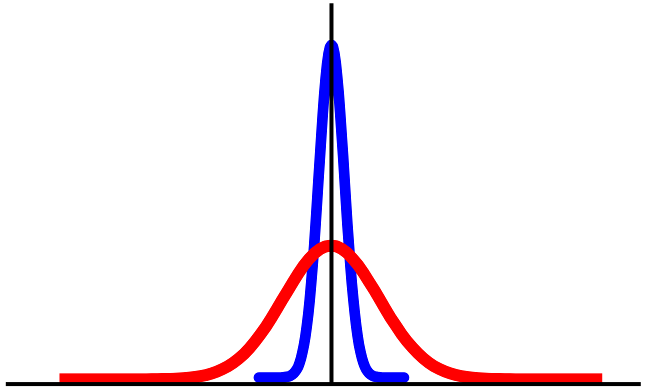


Normal Probability Plots

- Probability Distributions

Normal Distribution Other Characteristics:

- **Skewness:**
  - It is a measure of departure from symmetry.
  - Distributions may be skewed to the right (positive skew) or to the left (negative skew).
- **Kurtosis:**
  - It is a measure of the extent to which a distribution is flat-topped or peaked.
  - A positive value indicates a more peaked distribution than Normal.



- Probability Distributions

Further Considerations:

- **Left skewed distributions** may occur where the process has a physical upper limit or only has an upper specification limit.
- **Example:** the arrival time of job application since most job vacancies have a deadline and most applications arrive just before the deadline.
- **Right skewed distributions** are typically found when measuring time.
- This is because there is usually natural lower limit to how fact a process can be completed.
- **Examples:** The time for processing a job application.

- Probability Distributions

Further Considerations:

- It is important to understand the Normal distribution so that tools and techniques can be applied in a statistical valid manner.
- **Where the Normal distribution is important:**
  - **In Capability Analysis:** to make prediction about the process capability over the long term.
  - **In SPC:** certain charts require the data to be Normally distributed.
  - **In Hypothesis Testing:** many tests are based on the assumption that the data is Normally distributed.
  - **In Regression:** if the residual errors are Normally distributed, this indicates the regression model fits the data well.
  - **In Confidence Intervals and Design of Experiments.**