

# Graphical and Measurement Tools

# User Agreement and Copyright Information

- This recording and the accompanying guide contain copyrighted and proprietary content of Air Academy Associates, LLC. You are authorized to use this material for personal reference, but not for any commercial use. You may not modify, license, sub-license, distribute, copy, translate or create derivative works based on this guide, in part or in whole, without permission from Air Academy Associates.
- Other copyright information:
  - Six Sigma is a service mark of Motorola, Inc. Microsoft® and Excel® are registered trademarks of Microsoft Corporation in the United States and in other territories.
  - SPC XL™ and DOE Pro XL™ are copyright SigmaZone.com and Air Academy Associates, LLC. You may not copy, modify, distribute, display, license, reproduce, sell or use commercially any screen shots or any component contained therein without the express written permission of SigmaZone.com and Air Academy Associates, LLC. All rights reserved. SigmaZone.com may be contacted at [www.SigmaZone.com](http://www.SigmaZone.com). Air Academy Associates may be contacted at [www.airacad.com](http://www.airacad.com).

# Graphical and Measurement Tools

- In this session, we will discuss:
  - Why Collect Data
    - Attribute and Variables Data
  - Graphical Presentation
    - Pareto Chart – Attribute data
    - Histogram – Variables data
    - Run Chart – Both types
    - Box Plots – Both types
    - Scatter plot – Typically variables data
  - Measures (metrics) Presentation
    - Variables data – Summary statistics
    - Attribute data – Yield, first pass yield (FPY), rolled throughput yield (RTY), and defects per unit (dpu)
    - Variables data – Sigma level,  $C_p$ ,  $C_{pk}$ , and defects per million (dpm)
  - Summary
- A list of supplemental material and additional practice/review questions for this session are provided at the end of this presentation
- You can download the pdf of this presentation, along with any supporting data files, on the site where you are accessing this course

Take  
Note

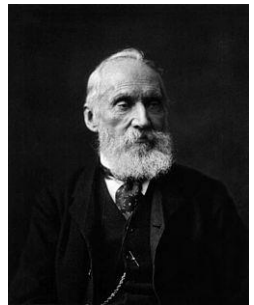
# Importance of Measurement

- Why collect data in Lean Six Sigma (LSS) and Design for Six Sigma (DFSS)?
  - To gather the facts for good decision making
  - Because perception and intuition are not always reality
  - To identify/verify problem areas and bottlenecks
  - To characterize our processes
  - To baseline the process
  - As an indicator of process performance

“To *measure* it, is to *know* it.”

“If you cannot *measure* it, you cannot *improve* it.”

“When you can *measure* what you are speaking about, and express it in numbers, you *know* something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meager and unsatisfactory kind; it may be the beginning of knowledge, but you have scarcely, in your thoughts, advanced to the stage of *science*.”



Lord Kelvin  
1824 - 1907



# Attribute and Variables Data

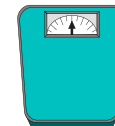
- **Attribute Data**

- Also called *categorical* data
- No direct measurement is made
- The presence or absence of something is recorded
- We classify or count a feature or characteristic of the product or process
- Examples:
  - Documentation errors per week
  - Test results (pass/fail, good/bad) out of an audit of 20 items
  - Presence or Absence of a date code stamp per product type



- **Variables Data**


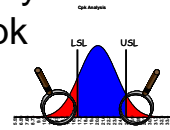
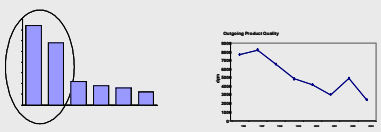
- Also called *continuous* or *measurement* data
- Actual measured values are recorded
- Data can assume a range of values on a continuous scale
- Examples:
  - Wait time in hours
  - Cost in dollars
  - Temperatures in degrees Fahrenheit or Celsius



- We will discuss methods to deal with both types of data, because we will likely experience both in our LSS and DFSS work.
- Given the choice, which type of data do you think is preferred and why?

# Measuring the Process

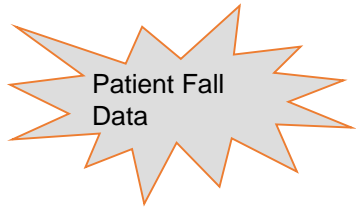
## Type of Summary

		Type of Summary	
		Graph	Metric, Measure
Type of Data	Variables (Measured)	<ul style="list-style-type: none"> <li>• Histogram</li> <li>• Run Chart</li> <li>• Box Plot</li> <li>• Scatter Plot</li> <li>• Control Chart</li> </ul> 	<ul style="list-style-type: none"> <li>• Mean, Standard deviation</li> <li>• Defects per million (dpm)</li> <li>• Sigma level</li> <li>• Sigma capability</li> <li>• Cp and / or Cpk</li> </ul> 
	Attribute (Counted)	<ul style="list-style-type: none"> <li>• Pareto Diagram</li> <li>• Run Chart</li> <li>• Box Plot</li> <li>• Control Chart</li> </ul> 	<ul style="list-style-type: none"> <li>• First pass yield (FPY)</li> <li>• % "good" (or "bad" )</li> <li>• Defects per unit (dpu)</li> <li>• Defects per million opportunities (dpmo)</li> <li>• Sigma capability</li> <li>• Defects per million (dpm)</li> </ul>

Goal: Turn data into information



**Graphical and  
Measurement Tools –  
data files**



**Data**

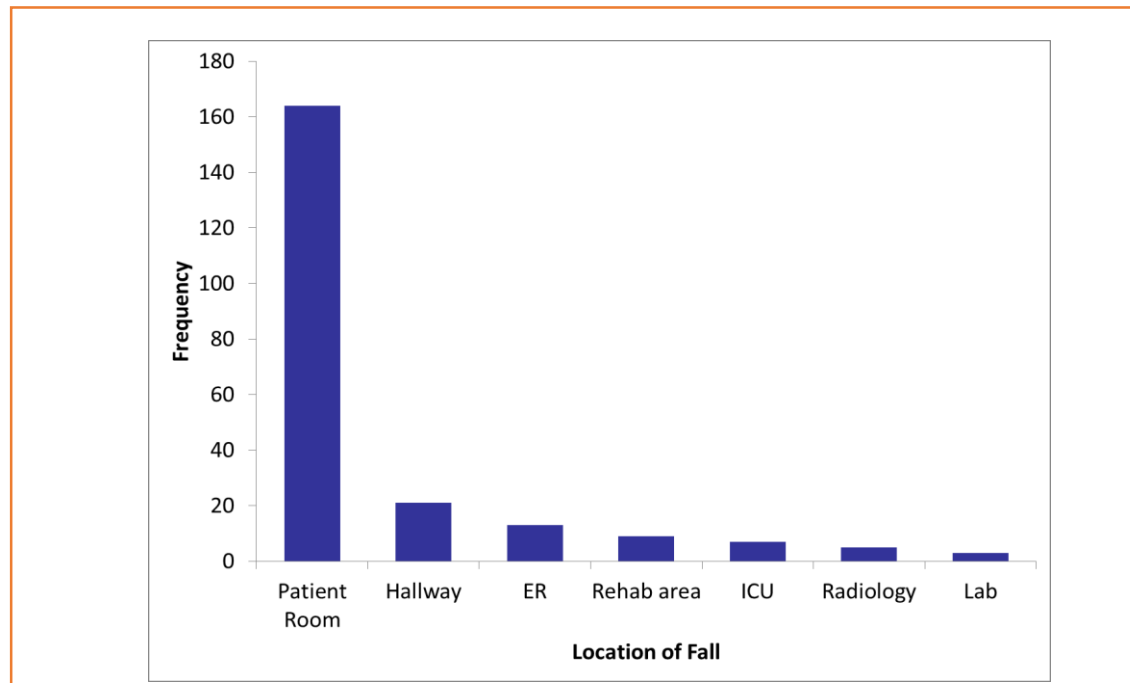
# Attribute Data: Pareto Example

## Patient Fall Location Data

Location of Patient Falls	Count (frequency)	Proportion
Hallway	21	0.09
Patient Room	164	0.74
Radiology	5	0.02
ER	13	0.06
Rehab area	9	0.04
ICU	7	0.03
Lab	3	0.01
	<u>222</u>	<u>1.00</u>

**Location of Falls and Frequency of Occurrence**

## Information

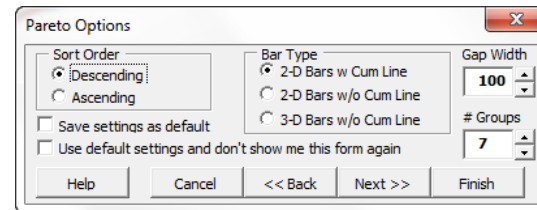
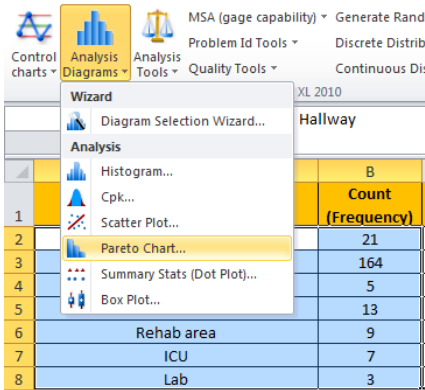


# Creating Pareto Charts with SPC XL

## SPC XL > Analysis Diagrams > Pareto Chart

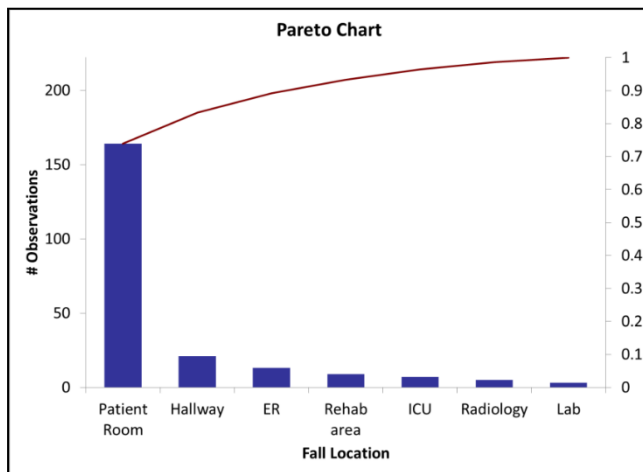
For video instruction on generating pareto charts, go to:

<https://airacad.com/our-insights/training-videos/spc-xl/>



Options window for chart details

Pareto Chart with a cumulative line shown

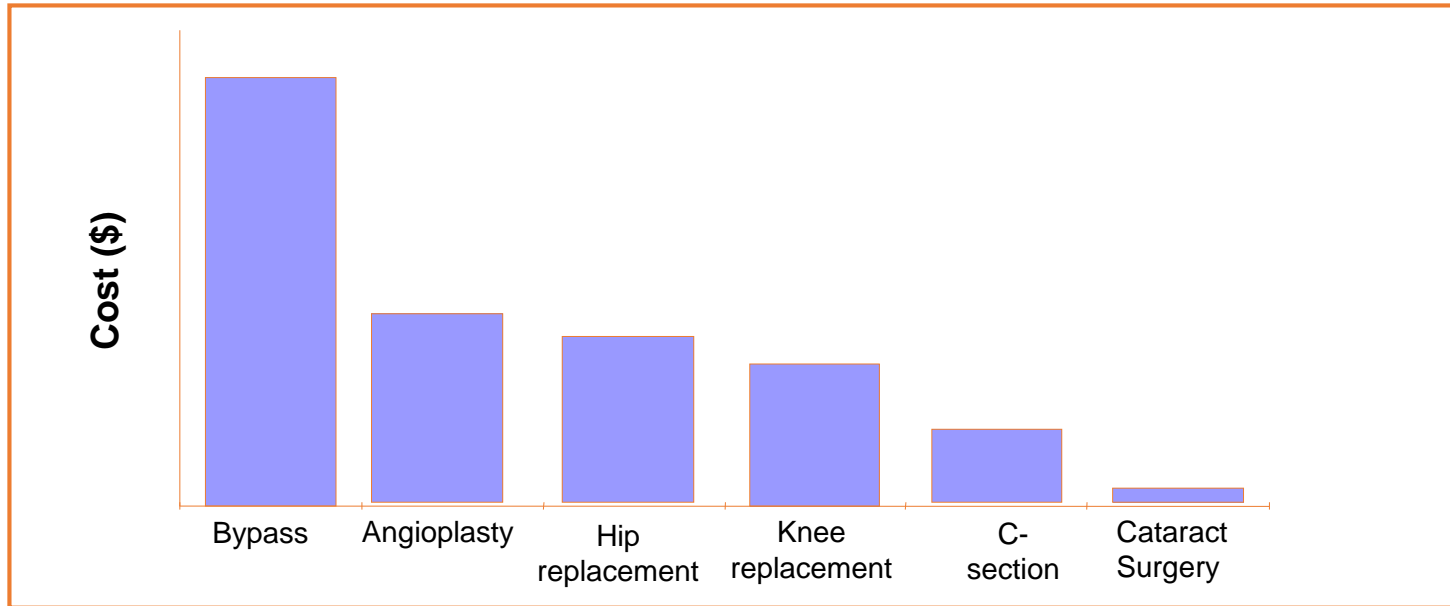


A Pareto Report is produced along with the Pareto Chart

Group	Count	% of Total
Patient Room	164	73.9%
Hallway	21	9.5%
ER	13	5.9%
Rehab area	9	4.1%
ICU	7	3.2%
Radiology	5	2.3%
Lab	3	1.4%



# Other Examples of Pareto Charts



**Pareto of Procedure Type by Cost**



**Pareto of Surgical Setup Errors by Time Lost Due to the Error**

# Variables Data: Histogram Example



**Graphical and Measurement Tools – data files**

n = 50

DATA

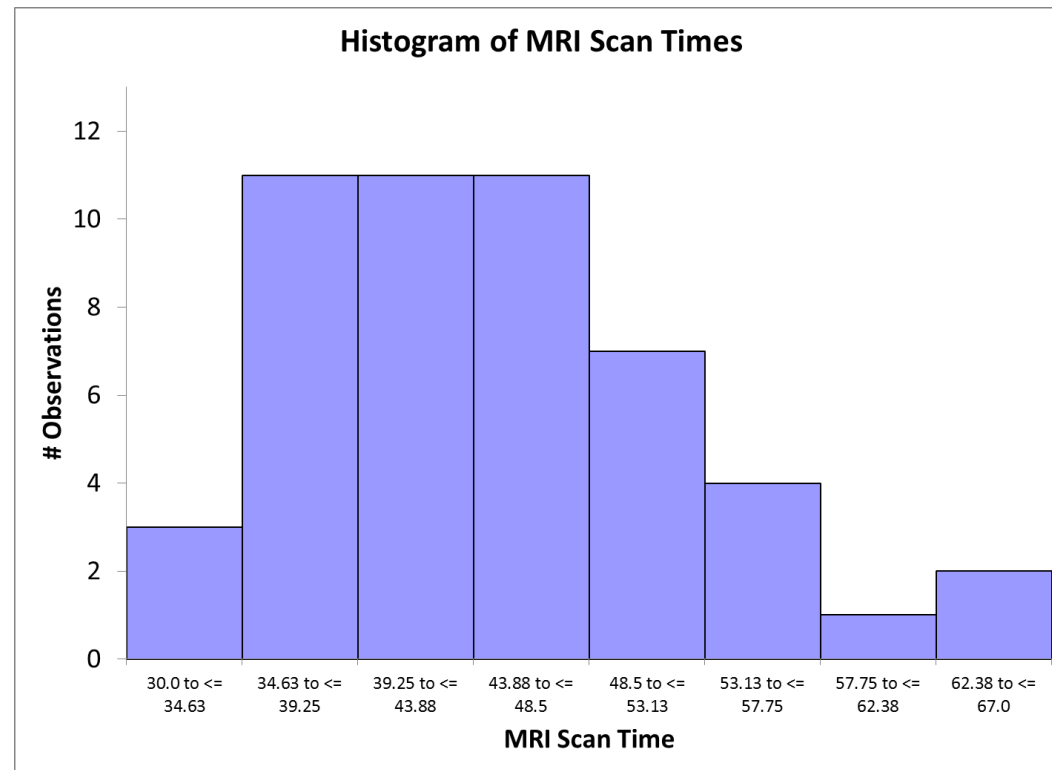
**Data**

40	38	52	51	50	43	39	63	30	39
54	48	38	46	49	39	39	55	41	40
53	49	45	60	55	36	35	48	33	50
43	41	47	44	39	34	43	43	47	48
45	42	44	41	43	36	67	37	46	56



## Information:

Histogram describes the distribution of the data: shape, location and spread!

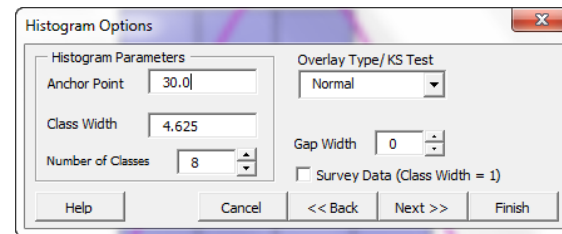
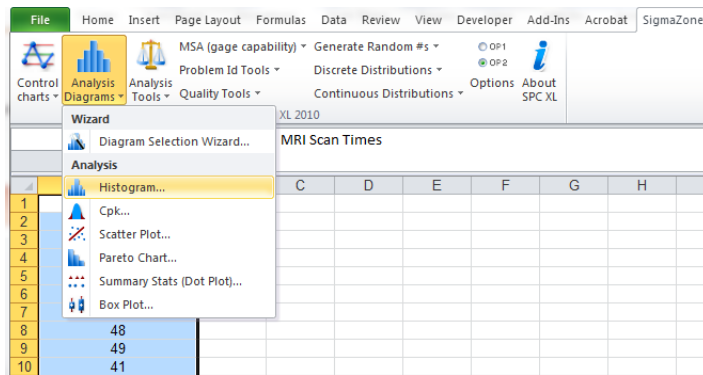


**Histogram of MRI Scan Times, for a particular type of scan**

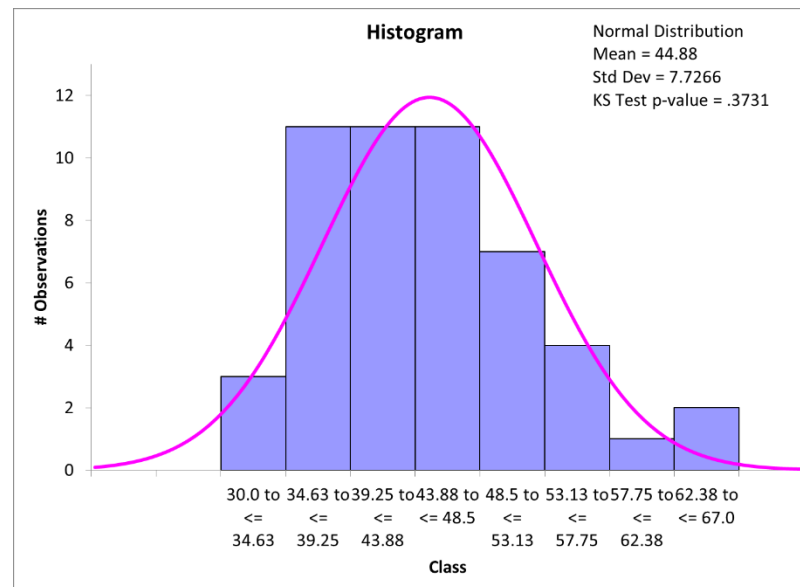
# Creating Histograms with SPC XL

For video instruction on generating histogram charts, go to:  
<https://airacad.com/our-insights/training-videos/spc-xl/>

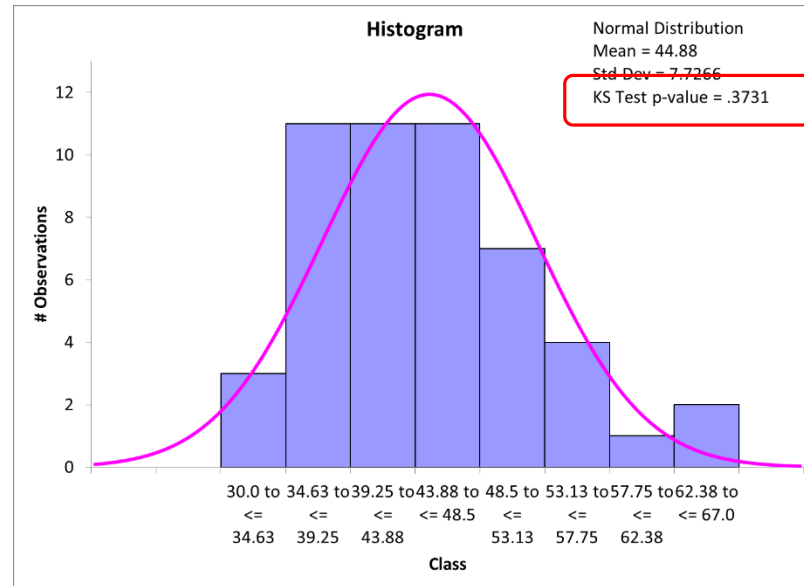
## SPC XL > Analysis Diagrams > Histogram



Options window for chart details



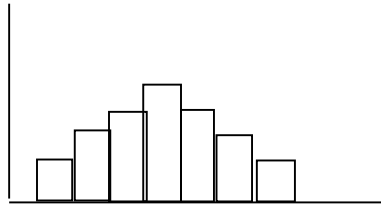
# Goodness of Fit Test



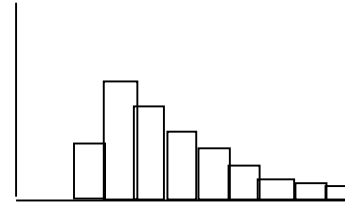
- The Kolmogorov-Smirnov (KS) Test is a “goodness of fit” test. The test result is displayed as a p-value (probability value).
  - $H_0$ : The distribution is normally distributed
  - $H_1$ : The distribution is not normally distributed
- Rule of Thumb:
  - If the p-value < 0.05, then conclude that the normal distribution IS NOT a good fit for your data ( $H_1$ ).
  - If p-value > 0.05, then we fail to reject  $H_0$  (the data could be normally distributed).

# Common Histogram Shapes

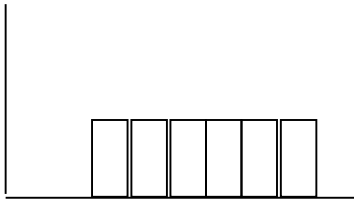
Classic “bell-shaped” curve



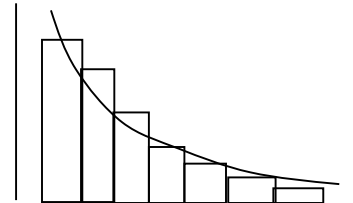
Skewed distribution



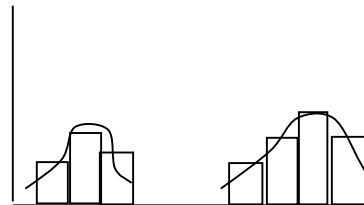
Rectangular (uniform) distribution



Exponential distribution



Bi-modal distribution



# Attribute and Variables Data: Run Chart

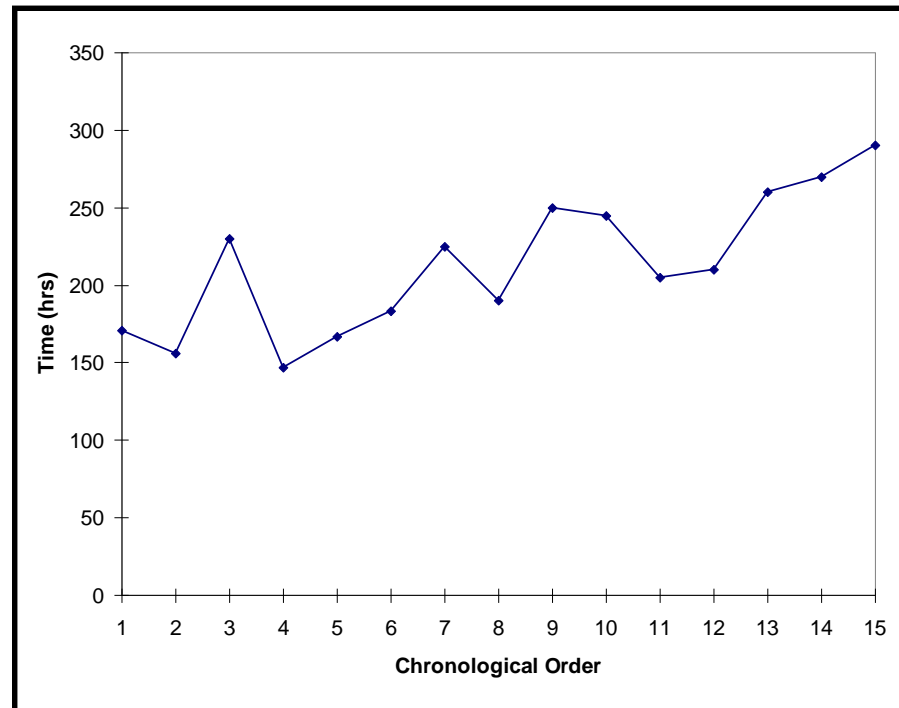
**Data:** 171, 156, 230, 147, 167, 183, 225, 190, 250, 245, 205, 210, 260, 270, 290

Time between machine breakdowns



**Graphical and Measurement Tools – data files**

## Information

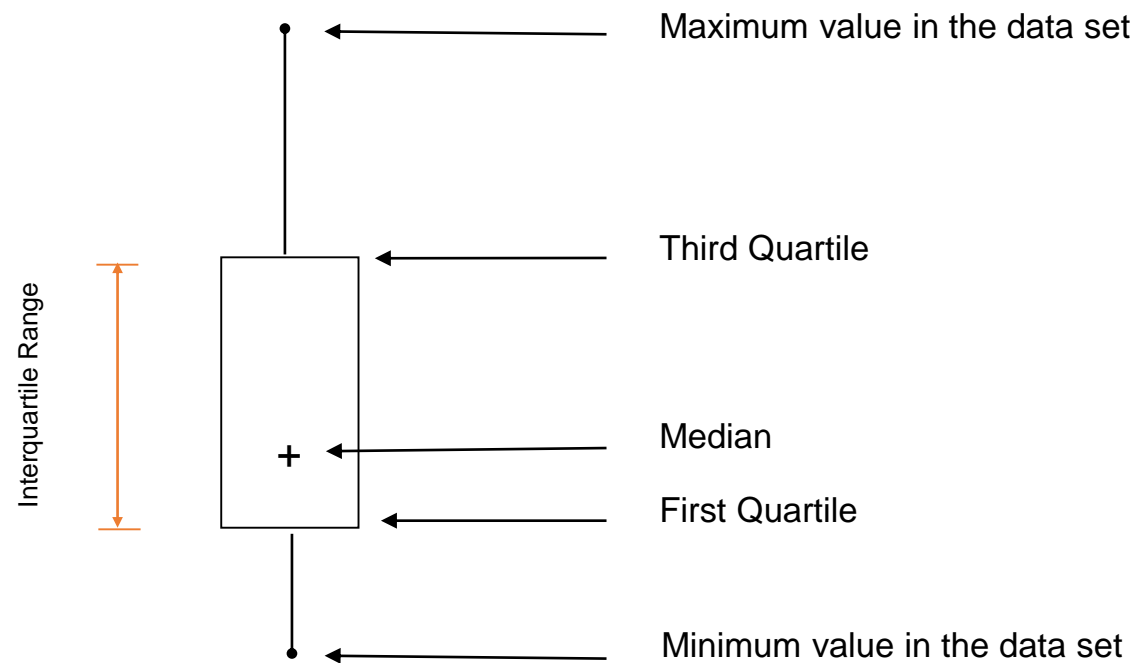


Run Chart for Time Between Breakdowns on a machine



# Box Plots

- Another method for graphically depicting variables data.
- Especially useful when comparing 2 or more sets of data (side by side) or when the data sets are small.
- Breaks up data into quartiles (fourths).
- Box plots show the data set's location and spread but not really the shape of the data!

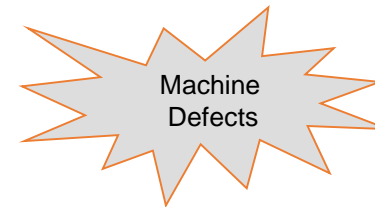
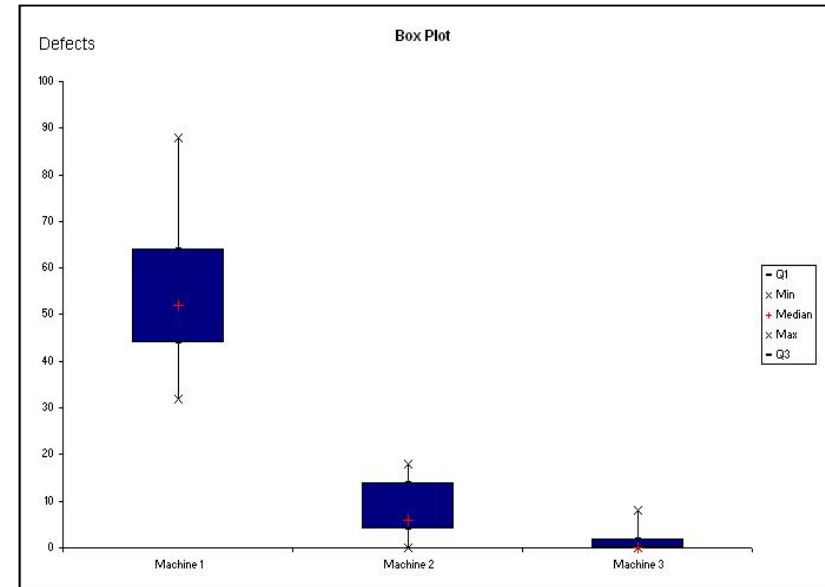
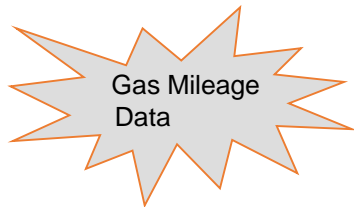
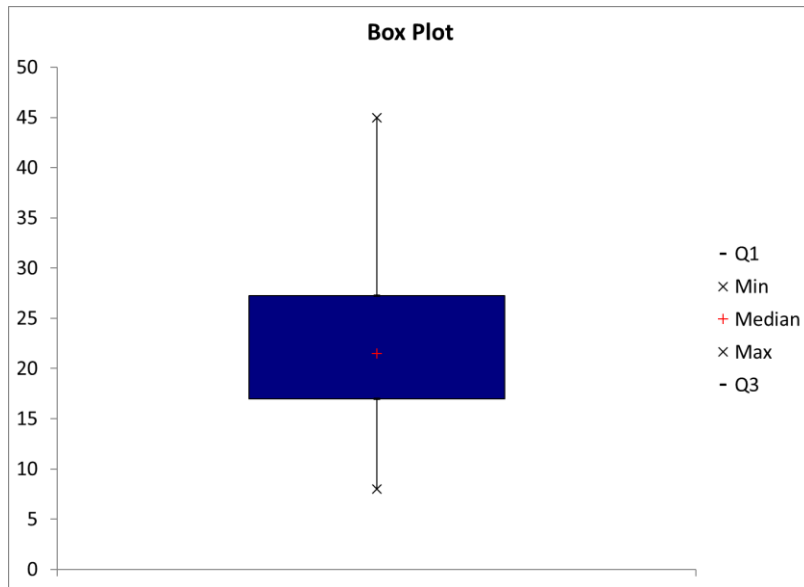




# Creating Box Plots with SPC XL

For video instruction on generating box plots, go to:  
<https://airacad.com/our-insights/training-videos/spc-xl/>

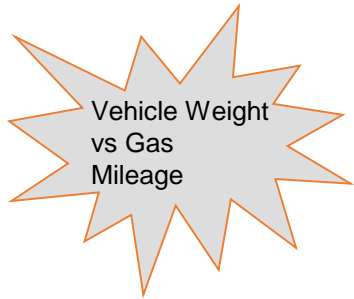
## SPC XL > Analysis Diagrams > Box Plot



**Graphical and Measurement Tools – data files**

# Scatter Plot Example

Data: Bivariate Data (Vehicle Weight vs Gas Mileage)

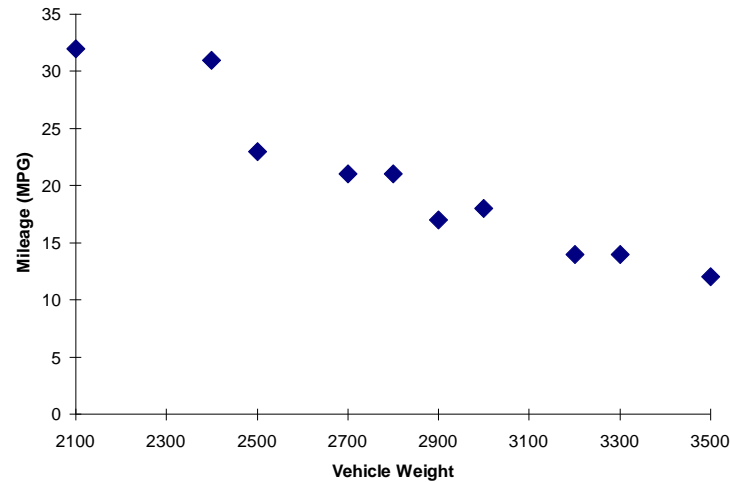


Observation	Weight (lbs)	Mileage (mpg)
1	3000	18
2	2800	21
3	2100	32
4	2900	17
5	2400	31
6	3300	14
7	2700	21
8	3500	12
9	2500	23
10	3200	14



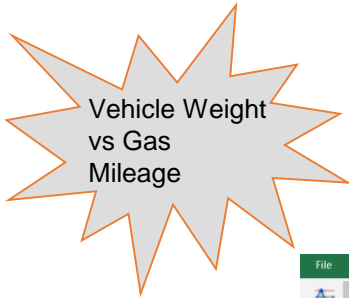
**Graphical and Measurement Tools – data files**

## Information

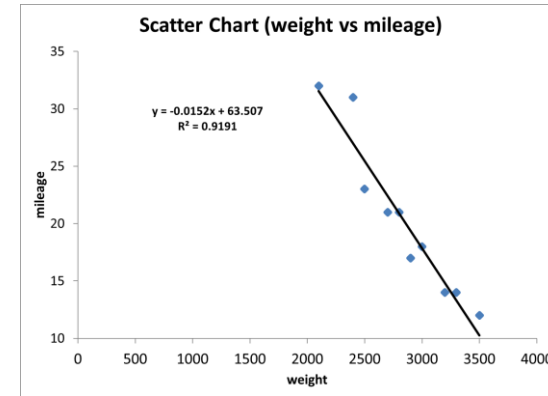
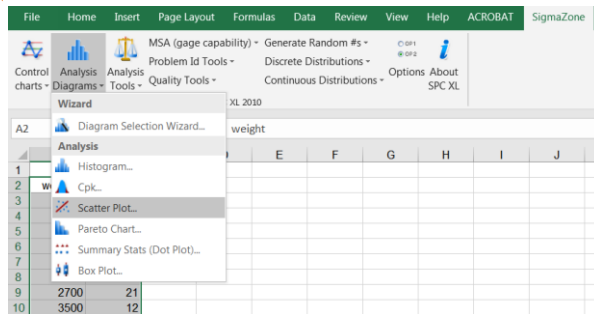


# Creating Scatter Plots with SPC XL

For video instruction on generating scatter plots, go to:  
<https://airacad.com/our-insights/training-videos/spc-xl/>

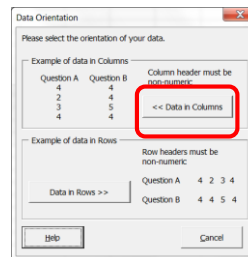
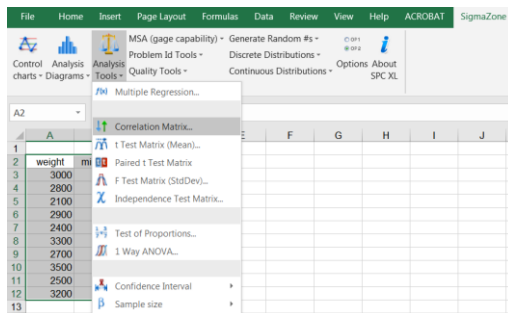


## SPC XL > Analysis Diagrams > Scatter Plot



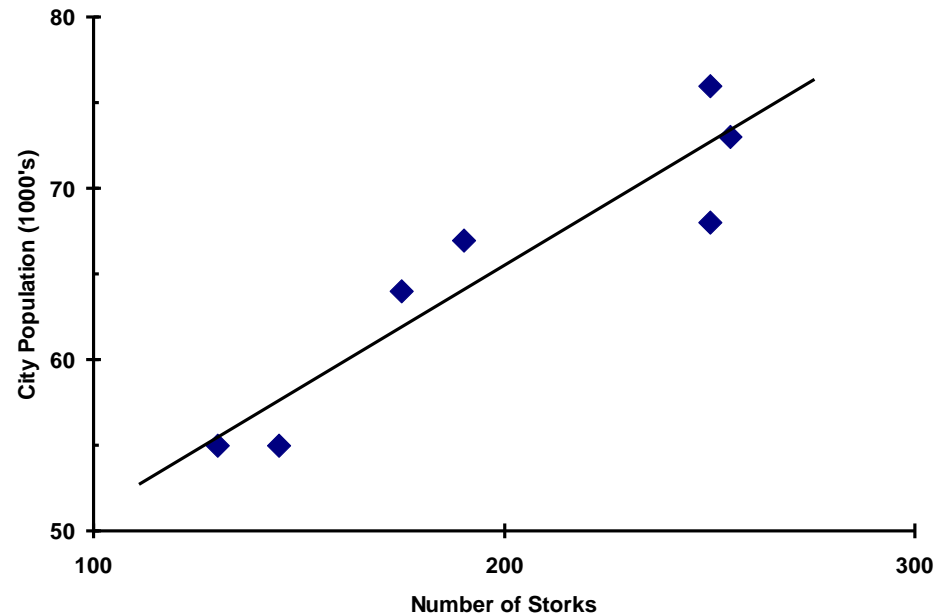
## SPC XL > Analysis Tools > Correlation Matrix

The table shown below displays the pairwise correlation coefficient.



Correlation Matrix		
	weight	mileage
weight	1.0	-0.958681
mileage		1.0
Summary		
Mean	2840.0	20.3
StDev	432.563	6.8646
Count	10	10

# Correlation vs Causality


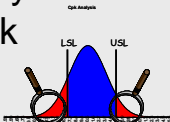
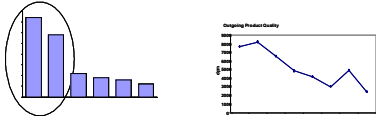


A plot of the population of Oldenburg, Germany at the end of each year against the number of storks observed in that year, 1930-1936.

*Source: "Statistics for Experimenters" by Box, Hunter, and Hunter. (1978)*

# Measuring the Process

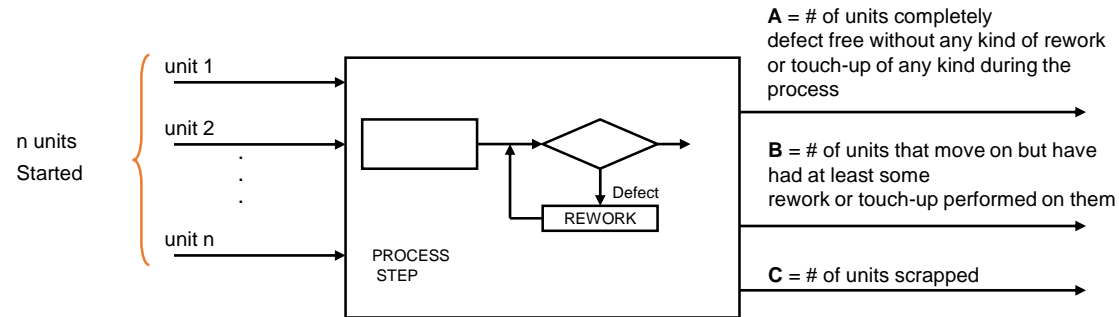
## Type of Summary

		Type of Summary	
		Graph	Metric, Measure
Type of Data	Variables (Measured)	<ul style="list-style-type: none"> <li>• Histogram</li> <li>• Run Chart</li> <li>• Box Plot</li> <li>• Scatter Plot</li> <li>• Control Chart</li> </ul> 	<ul style="list-style-type: none"> <li>• Mean, Standard deviation</li> <li>• Defects per million (dpm)</li> <li>• Sigma level</li> <li>• Sigma capability</li> <li>• Cp and / or Cpk</li> </ul> 
	Attribute (Counted)	<ul style="list-style-type: none"> <li>• Pareto Diagram</li> <li>• Run Chart</li> <li>• Box Plot</li> <li>• Control Chart</li> </ul> 	<ul style="list-style-type: none"> <li>• First pass yield (FPY)</li> <li>• % "good" (or "bad" )</li> <li>• Defects per unit (dpu)</li> <li>• Defects per million opportunities (dpmo)</li> <li>• Sigma capability</li> <li>• Defects per million (dpm)</li> </ul>

Goal: Turn data into information

# Summarizing the Current State with Measures of Quality (Attribute Data)

Unit: any part, item, component, subassembly or complete product, service, or transaction for which a quality/performance measure is desired. This must be specifically defined.

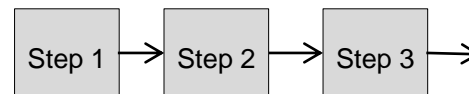


$$\text{Yield} = \frac{A + B}{n}$$

$$\text{First Pass Yield (FPY)} = \frac{A}{n}$$

Rolled Throughput Yield (RTY)

$$\text{RTY} = \text{FPY1} \times \text{FPY2} \times \text{FPY3}$$



$$\text{defects per unit (dpu)} = \frac{\text{Total \# of defects observed from start to finish over all units started}}{\text{Total \# of units started (n)}}$$

# Measuring Quality for Attribute Data with SPC XL

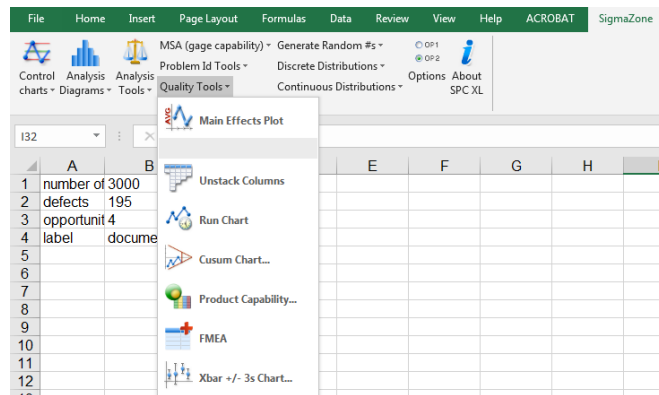
## SPC XL > Quality Tools > Product Capability



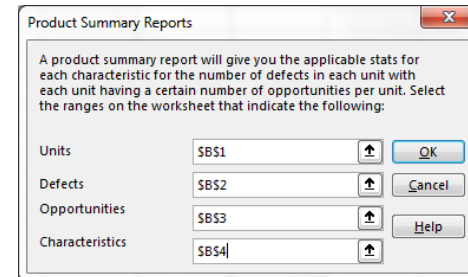
- Record pertinent information in your spreadsheet:
  - Number of units tested or inspected
  - Number of defects found
  - Number of opportunities per unit (optional)
  - Text description (optional)



**Graphical and Measurement Tools – data files**



Reference the cells where your data is stored

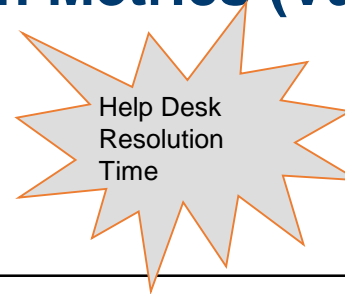


Product Summary Report								
Characteristic	Defects	Units	Opportunities	Total Opportunities	DPU	DPO	DPMO	Sigma Capability
documents	195	3000	4	12000	0.06500	0.01625	16250.000	3.6382
<b>Total</b>	<b>195</b>		<b>0</b>	<b>12000</b>		<b>0.01625</b>	<b>16250.000</b>	<b>3.6382</b>

# Summarizing the Current State with Metrics (Variables Data)



**Graphical and  
Measurement Tools –  
data files**



6	7	7	9	11	12	15
---	---	---	---	----	----	----

## Mean:

(measure of location and center;  
average; balance point)

$$\bar{y} = \frac{\sum_{i=1}^n y_i}{n}$$

$$\bar{y} = \frac{6+7+7+9+11+12+15}{7} = 9.57$$

## Median:

(another measure of center; used  
with highly skewed data; 50<sup>th</sup> percentile)

Median = middle value  
= 9

## Process standard deviation:

(spread; variation)

$$s = \sqrt{\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-1}}$$

$$= \sqrt{\frac{(6-9.57)^2 + (7-9.57)^2 + (7-9.57)^2 + (9-9.57)^2 + (11-9.57)^2 + (12-9.57)^2 + (15-9.57)^2}{6}}$$
$$= 3.26$$



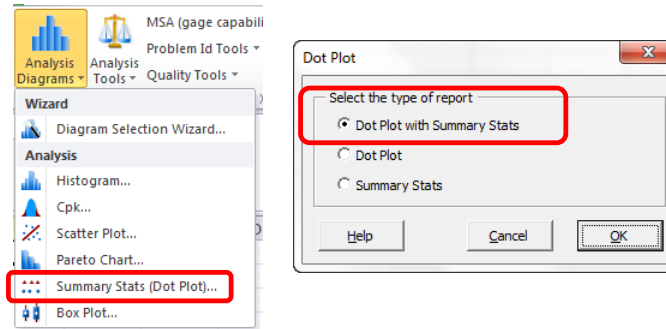
# Summary Statistics with SPC XL

- For video instruction on generating summary statistics, go to: <https://airacad.com/our-insights/training-videos/spc-xl/>



**Graphical and Measurement Tools – data files**

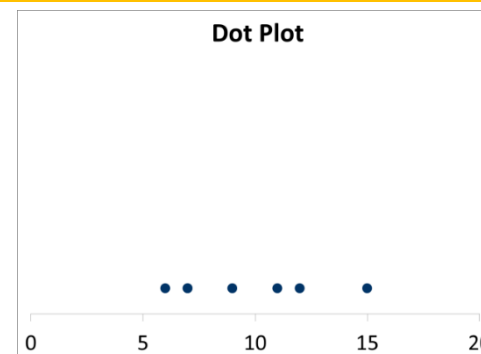
## SPC XL / Analysis Diagrams / Summary Stats



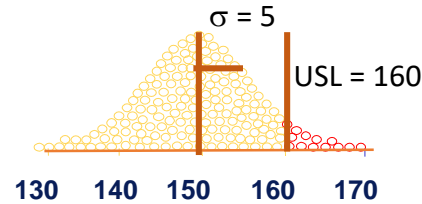
Count	7
Mean	9.571428571
Median	9
Mode	7
Max	15
Min	6
Range	9
Std Dev (Pop)	3.016958869
Std Dev (Sample)	3.258688021
Variance (Pop)	9.102040816
Variance (Sample)	10.61904762
95% Conf. Interval for Mean	
Upper Limit	12.58521233
Lower Limit	6.55764481
99% Conf. Interval for Mean	
Upper Limit	14.13775014
Lower Limit	5.005107001

The summary statistics which SPC XL displays can be changed in “Options” “Summary Stats”.

The dot plot shows where the data falls and can be useful for identifying outliers.



# Summarizing the Current State with Measures of Quality (Variables Data)



- **Sigma Level** = number of standard deviations between the center of the process and the nearest specification limit  
 = minimum  $\left( \frac{USL - \bar{y}}{\sigma}, \frac{\bar{y} - LSL}{\sigma} \right)$
- **C<sub>p</sub>** = a process potential index  
 =  $\frac{\text{specification width}}{\text{process width}}$
- **C<sub>pk</sub>** = a process capability index (actual)  
 = proportion of natural tolerances ( $3\sigma$ ) between the center of a process and the nearest spec  
 = minimum  $\left( \frac{USL - \bar{y}}{3\sigma}, \frac{\bar{y} - LSL}{3\sigma} \right) = \frac{\sigma_{level}}{3}$
- **Defects per million (dpm):** The number of times (out of a million) that the process doesn't meet the specifications (requirements) (e.g., 20% is equivalent to 200,000 per million)

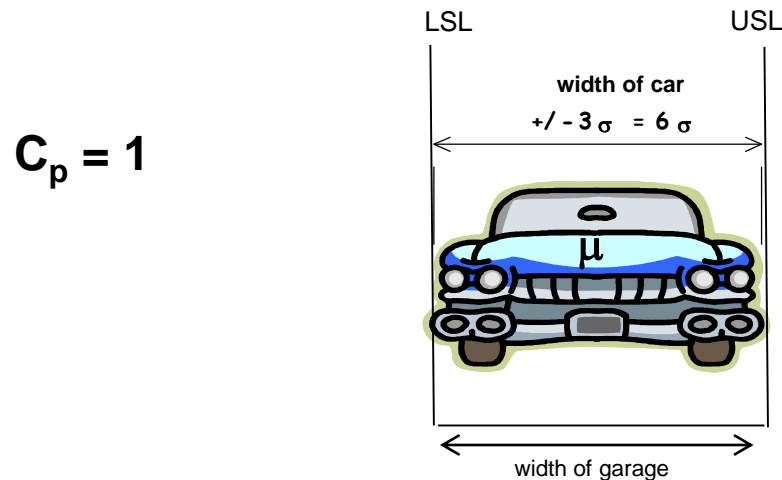
Note: The metrics Cp and Cpk assume that data is normally distributed

# Cp Review (Variables Data)

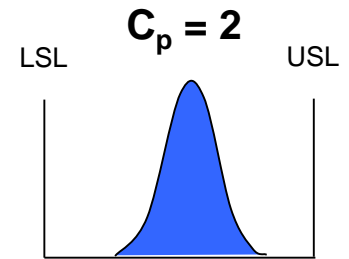
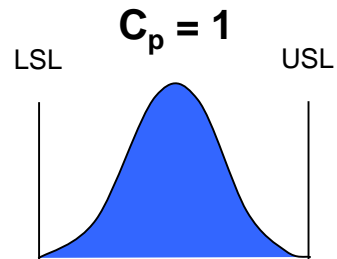
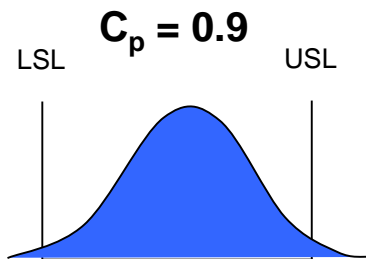
$C_p$ : Process Capability Potential (assumes a centered process)

$$C_p = \frac{USL - LSL}{6\sigma} = \frac{\text{voice of the customer}}{\text{voice of the process}}$$

Interpreting the  $C_p$  value (“garage and the car”)



Examples:



# Cpk Review (Variables Data)

$C_{pk}$ : Process Capability (Actual) (takes into consideration a shift in the average)

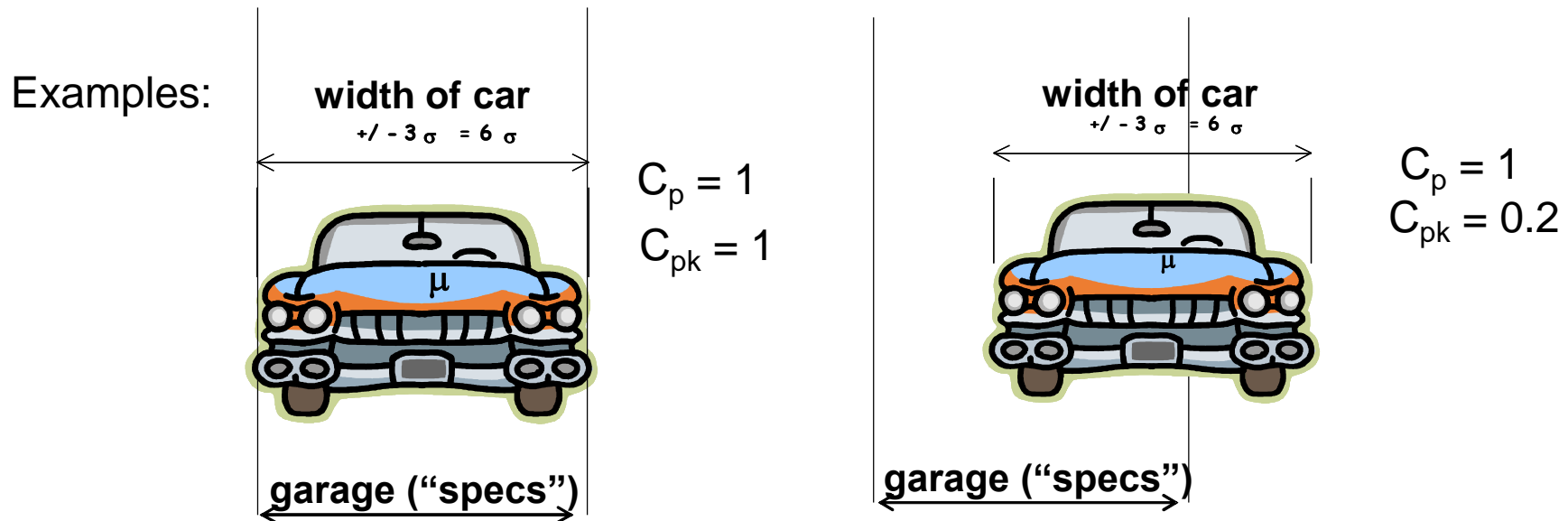
$$C_{pk} = \min\left(\frac{USL - \bar{y}}{3\sigma} \text{ or } \frac{\bar{y} - LSL}{3\sigma}\right) = \frac{\sigma_{level}}{3}$$

$C_{pk} = C_p$  when the process is perfectly centered

$C_{pk} < C_p$  when the process is not centered

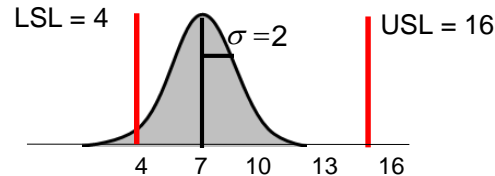
$C_{pk} = 0$  when the process mean is equal to one of the specification limits

$C_{pk}$  becomes negative when the process mean lies outside a specification limit



# Capability Analysis with SPC XL

Suppose the time to process a claim follows a normal distribution with a mean of 7 days and a standard deviation of 2 days. The upper spec limit is 16 days, with a lower spec limit (imposed by management) of 4 days.

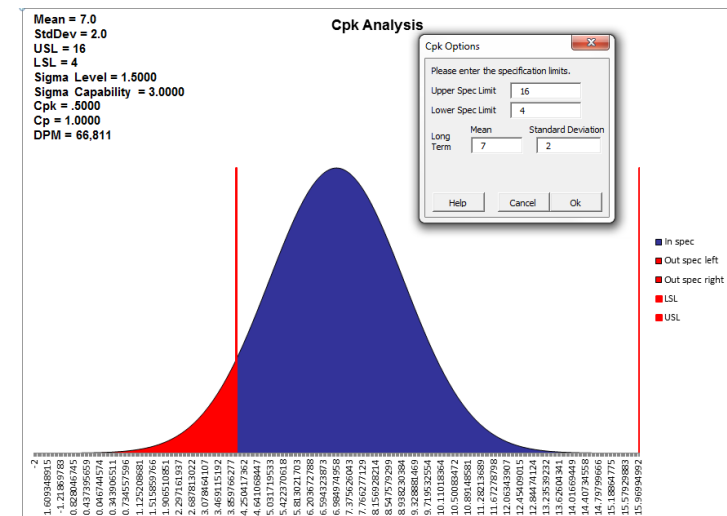
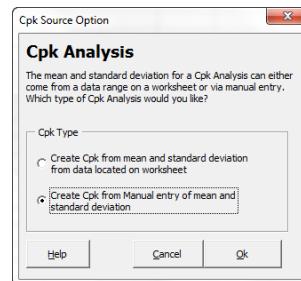
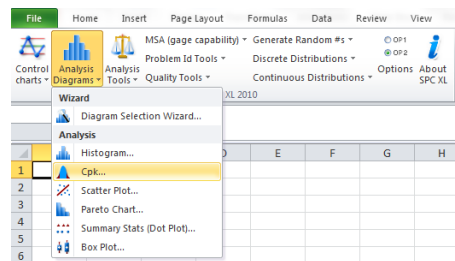


- For video instruction on generating a capability analysis, go to: <https://airacad.com/our-insights/training-videos/spc-xl/>


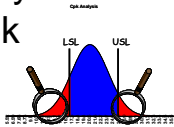
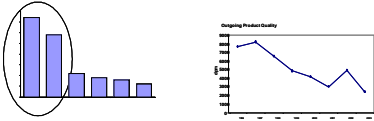
## SPC XL > Analysis Diagrams > Cpk

Input specs, mean, and standard deviation as shown

Source data can be stored in a spreadsheet, or you can manually enter the mean and standard deviation.



# Graphical and Measurement Tools Summary

		Type of Summary	
		Graph	Metric, Measure
Type of Data	Variables (Measured)	<ul style="list-style-type: none"> <li>Histogram</li> <li>Run Chart</li> <li>Box Plot</li> <li>Scatter Plot</li> <li>Control Chart</li> </ul> 	<ul style="list-style-type: none"> <li>Mean, Standard deviation</li> <li>Defects per million (dpm)</li> <li>Sigma level</li> <li>Sigma capability</li> <li>Cp and / or Cpk</li> </ul> 
	Attribute (Counted)	<ul style="list-style-type: none"> <li>Pareto Diagram</li> <li>Run Chart</li> <li>Box Plot</li> <li>Control Chart</li> </ul> 	<ul style="list-style-type: none"> <li>First pass yield (FPY)</li> <li>% "good" (or "bad" )</li> <li>Defects per unit (dpu)</li> <li>Defects per million opportunities (dpmo)</li> <li>Sigma capability</li> <li>Defects per million (dpm)</li> </ul>

Goal: Turn data into information

# Key Takeaways



- As a review techniques, stop the video and summarize the key learnings from this session. When you are finished, continue to the next page.

# Key Takeaways

- **Why collect data?**
  - For many reasons, but primarily to baseline current performance
- **Types of Data**
  - Attribute – counted, Variables – measured
- **Graphical Presentation – a picture!!!**
  - Pareto Chart – attribute data (separates the vital few from the trivial many!)
  - Histogram – variables data (provides a distribution – shape, location, and spread)
  - Run Chart – both types of data (data over time, shows trends and shifts)
  - Box Plot – both types of data (compares location and spread differences across multiple groups)
  - Scatter Plot – variables data (investigates relationships between two variables)
- **Measures (Metrics) Presentation – capability! (current performance)**
  - Variables Data – Summary statistics (Mean, Median, Standard Deviation)
  - Attribute Data – Yield, FPY, RTY, and dpu (quality metrics!)
  - Variables Data – Sigma level,  $C_p$ ,  $C_{pk}$ , and dpm (quality metrics!)



# Supplemental Material



- Suggested Reading:
  - ***Lean Six Sigma: A Tools Guide*** by Adams, Kiemele, Pollock and Quan (pp. 111 – 128, 223 - 227)
  - ***Basic Statistics – Tools for Continuous Improvement*** by Kiemele, Schmidt and Berdine, 4<sup>th</sup> edition (Chapter 2)
  - ***Design for Six Sigma: The Tool Guide for Practitioners*** by Reagan and Kiemele (Section 7.15)
  - Air Academy's app: ***Six Sigma Quick Tools***



- SPC XL™ software training tutorials:
  - <https://airacad.com/our-insights/training-videos/spc-xl/>
- The data files for this session can be downloaded from the site where you are accessing this course.

# Additional Practice / Review Questions



## **Graphical and Measurement Tools – data files**



- 1) In a review process, one of the critical measures of performance is the review cycle time. The goal is to complete the review below 72 hours. Baseline data was collected chronologically and is in the Review Cycle Time worksheet of the data files for this session. Display the data graphically and summarize the quality and capability of the process.
- 2) A customer service organization has been tracking the time its customers spend on hold along with the proportion of abandoned calls. Data for the past 15 days is located in the Hold Time Abandon Rate worksheet of the data files for this session. Is there a strong relationship between the average hold time and the abandon rate? Why or why not? Do you have any other observations or comments about the data?

# We can help...

## Connect With Us



### [Remote Project Coaching](#)

There are times when help outside your organization is needed. When that time comes, benefit from a partner that is experienced, tested, and trusted.

Expert coaching is one of the Top Five Best Practices for generating step change in project execution, as well as enhanced return on investment. We can work remotely with your organization to provide coaching support.

Air Academy Associates

Phone: (719) 531-0777

Email: [aaa@airacad.com](mailto:aaa@airacad.com)

<https://airacad.com/>

<https://sixsigmaproductsgroup.com/>



There's an app for that!  
***Six Sigma Quick Tools***

