Statistical Tools – Handy Guide- Performance Qualification

Definitions

SPC – Statistical Process Control

Specification Limits – Specific requirements for a product, material or process. For example, %Base Solids in a particular product: Min 9% max 11%

Control Limits – The limits between which you expect variability due to common cause (usually set at ±3 standard deviations from the mean)

Variability (or Variation) – Differences

KPC – Key Product Characteristic

Common Cause Variability – Inherent to the process, consistent and predictable effect on the variability of individual items

Special Cause Variability – sources of variability that are unique and lead to unacceptable variability.

Population – The whole of the data that we are interested in

Sample – a representative subset of the population

Factor – what you test to see if it affects a response (e.g. Brine, Citric)

Level – the specific conditions of a factor being tested under experimenter's control (e.g. concentration = 1% or 3%)

Treatment – a unique combination of levels of each factor (e.g. Brine = 1% with Citric = 3%)

Group – another term for a treatment

Response – thing you are measuring. Shows the effect of the factors (i.e., Viscosity = 2500 cps when Brine = 1% and Citric = 3%)

Replicate – when a treatment is tested 2 or more times

Mean – A measure of Average (total of all values divided by number of values)

Range – Difference between the maximum and minimum value

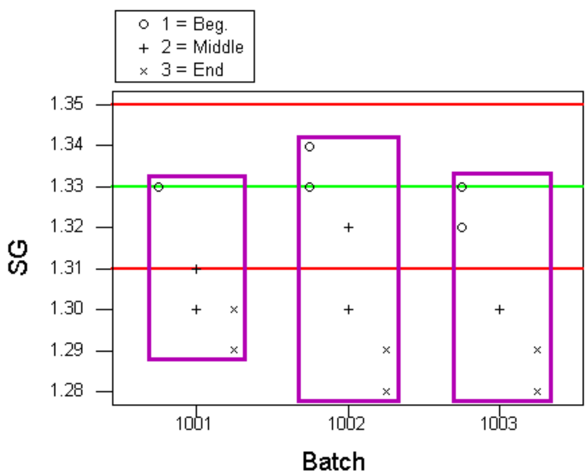
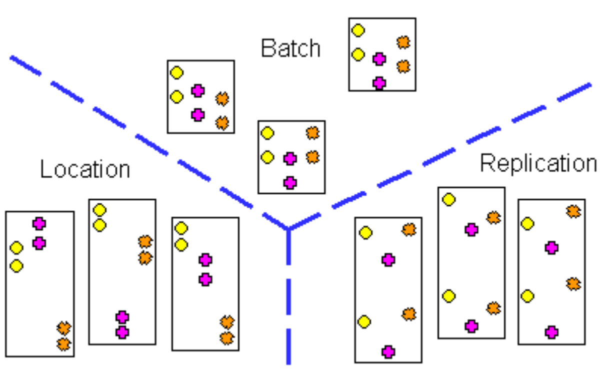
Standard Deviation – A measure of the variability of a process

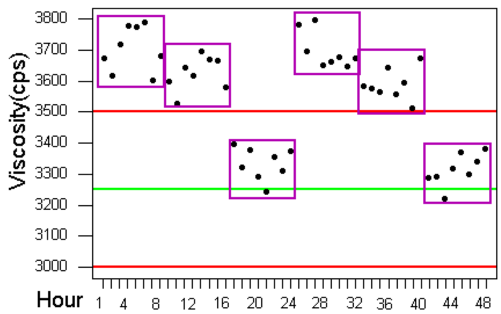
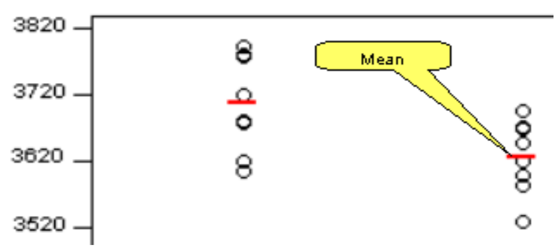
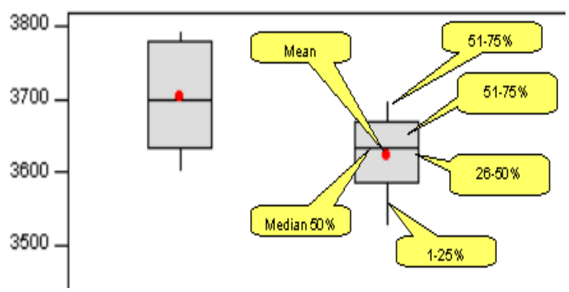
Variance – a measure of the variability of a process, the square of standard deviation

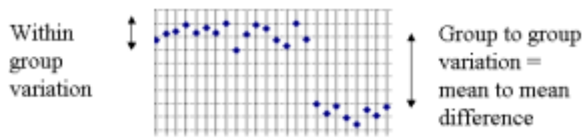
Normal Distribution – Pattern of variability which looks like a bell-shaped curve with Mean as the middle value of the curve

Random – Not influenced by the observer

Probability – the chance that something will or will not happen. Usually expressed as % or decimal (50% or 0.5)

Tool	Description	Use When	Use How
Snee Plot	 	1. To visually determine what greatest source of variability is: either Batch Variability, Location Variability, or Replication (Analysis or Method) Variability (see bottom graphic) 2. To notice non-random patterns. For example, in the top diagram the Specific Gravity is always highest for the beginning of the batch.	1. Sample each batch 6 times, two measurements (replications) at three sampling times (Beginning, Middle, End) or locations (Bottom, Middle, Top) 2. Plot data by hand or by computer. 3. Draw boxes by hand around data from each group (or by computer drawing tool) 4. Determine greatest source of variability by comparing plotted result with bottom graphic

Tool	Description	Use When	Use How
Boxes (for continuous process)		1. To gain more visual separation between <u>actual data</u> from different treatments. In pictured example, each box represents a different shift.	1. Sample at regular time intervals 2. Plot data by hand or by computer 3. Draw boxes around each group 4. Look for differences in group means
Dotplot and Boxplot		1. To see visual separation between <u>actual data</u> from different treatments in simple diagram 2. See variability unaffected by outliers (Boxplot) 3. See if data values are evenly distributed or if large proportion of data falls in certain data quadrants, i.e. above or below median (Boxplot)	1. Obtain data for a single variable where 2 or more treatments have been made that may influence your response 2. Create simultaneously with MINITAB ® * as an option while doing a One-Way ANOVA 3. Compare group means & medians & variability of groups by height

Tool	Description	Use When	Use How
ANOVA 1 Way (for one factor)	 **Accept or Reject Null Hypothesis** **Null Hypothesis - H_o** **No effect** **e.g. There is no difference among shifts** **Alternative Hypothesis - H_A** **There is an effect** **e.g. There is a difference among shifts** **F-ratio and p-ratio:** F = Between Groups / Within Groups p = probability of falsely rejecting the Null Hypothesis F large, p small → reject Null Hypothesis F small, p large → fail to reject Null Hypothesis **One-way ANOVA: Initial SG, PowderSolution SG** Analysis of Variance Source DF SS MS F P Factor 1 0.003969 0.003969 20.21 0.000 Error 34 0.006678 0.000196 Total 35 0.010647 Individual 95% CIs For Mean Based on Pooled StDev Level N Mean StDev ---------+---------+---------+------- Initial 18 1.3078 0.0190 (------*-----) PowderSo 18 1.3288 0.0058 (------*-----) ---------+---------+---------+------- Pooled StDev = 0.0140 1.310 1.320 1.330	1. To see if there is a high probability that one or more levels are affecting a response differently (as in different Mean) 2. You can assume that data is normally distributed and treatment variances are roughly equal. See a statistician if data is not normal.	1. Obtain data for a single variable where two or more different treatments have been made that may influence the variable 2. Make ANOVA table with MINITAB ® * 3. MiniTab ® * compares your F-value to F-distribution to give you a p-value, the probability of error you have in concluding that the groups are different (e.g. p = 0.05 is 5% prob. of error) 4. Compare p-value to required value. For example, p = 0.065 = 6.5 % but you want to be 95% confident (only 5% prob. of error) then you conclude groups are not different. 5. Confidence Interval diagram can be used along with p-value. Each confidence interval (CI) contains 95% of a population's expected values. When CIs are well separated & none of the means fall within another CI this supports conclusion that treatments do not produce identical response.

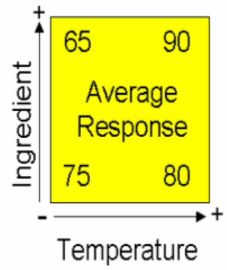
Tool	Description	Use When	Use How
Design of Experiments (2 X 2 shown) along with 2 Way ANOVA and Interaction Plot	(see below)	(see below)	(see below)

Description

Designed Experiment 1.MTW ***

↓	C1-T	C2-T	C3	C4	C5
	Std Order	Random Order	Temperature	Ingredient	Whiteness
1	1a	6	(-) 68	(-) 1	75
2	2a	5	(+) 110	(-) 1	82
3	3a	1	(-) 68	(+) 5	67
4	4a	8	(+) 110	(+) 5	91
5	1b	3	(-) 68	(-) 1	75
6	2b	4	(+) 110	(-) 1	78
7	3b	2	(-) 68	(+) 5	63
8	4b	7	(+) 110	(+) 5	89

(-) indicate "Low Level"   (+) indicates "High Level"



Data for a 2 X 2 Designed Experiment with Replication in Random Order

Two-way ANOVA: Whiteness versus Temp, Ingredient

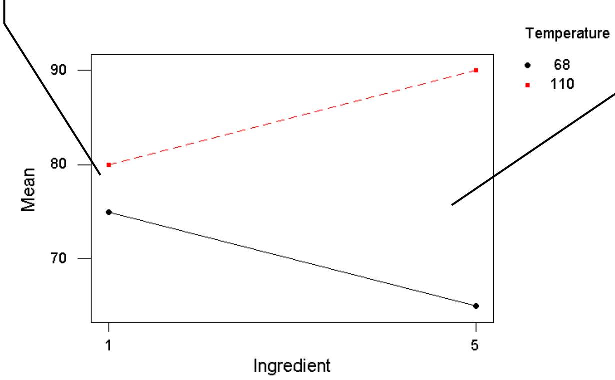
```
Analysis of Variance for Whitenes
Source        DF        SS        MS        F        P
Temp           1    450.00    450.00   100.00    0.001
Ingredie       1      0.00      0.00     0.00    1.000
Interaction    1    200.00    200.00    44.44    0.003
Error          4     18.00      4.50
Total          7    668.00

                        Individual 95% CI
Temp           Mean    ------+---------+---------+---------+----
 68            70.0    (-----*-----)
110            85.0                                   (-----*-----)
                       ------+---------+---------+---------+----
                          70.0      75.0      80.0      85.0

                        Individual 95% CI
Ingredie       Mean    ---+---------+---------+---------+--------
1              77.5    (-----------------*-----------------)
5              77.5    (-----------------*-----------------)
                       ---+---------+---------+---------+--------
                       75.0      76.5      78.0      79.5
```

The spreads between the Confidence Intervals seen show Temperature has significant effect, Ingredient does not

p-values show Temp and Interaction have effect, Ingredient does not

Non-parallel lines indicate an interaction among factors

Interaction Plot - Data Means for Whiteness



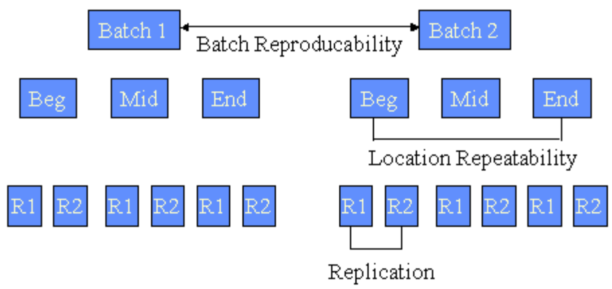
The spreads between the Interaction Plot Lines show that Temperature gives higher whiteness overall than Ingredient.

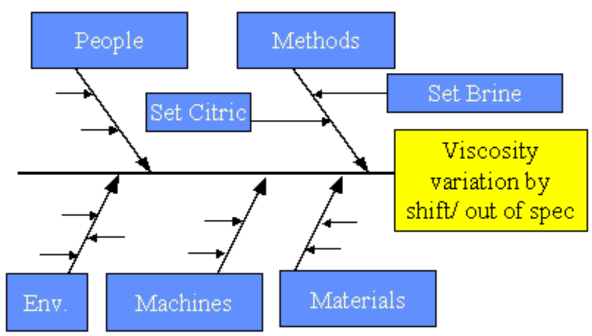
Use When

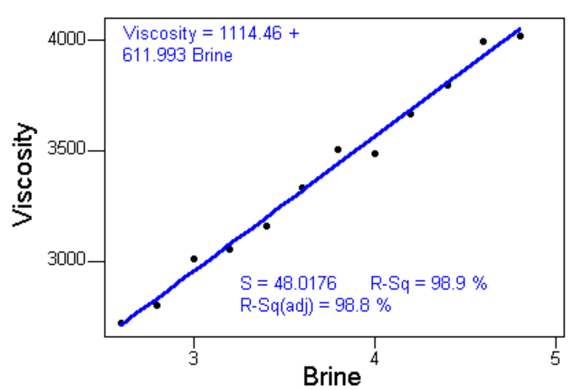
1. You have 2 formula or process factors that you can control that may influence a response and you want to see which has the greatest influence
2. You are also looking to see if the interaction among the 2 factors influences the response
3. You would like to optimize a response by using the best combination of factors which meet your production requirements and the customers requirements

Use How

1. Decide upon 2 factors that you believe to exert the strongest influence upon a response (e.g. Ingredient and Temperature influence Whiteness of laundry)
2. Choose 2 levels for each factor (68 and 110 for Temperature, 1% and 5% for Ingredient)
3. There are now four different treatments (e.g. T=68 with I = 5%, or T = 110 with I = 1%). To be able to measure an Interaction Effect though you must repeat each of these 4 treatment for 8 runs total.
4. Create a random order of the 8 runs to minimize the possibility of unknown factors confounding results (e.g. because of not occurring uniformly across different levels of the factors). See top diagram.
5. Run 2 Way ANOVA (middle diagram) and Interaction Plot (bottom diagram)
6. Analyze using p-values, Confidence Interval spread and Interaction Plot lines

<table>
<tr><th>Tool</th><th>Description</th><th>Use When</th><th>Use How</th></tr>
<tr>
<td>Components of Variance</td>
<td>


<table>
<tr><th>Source</th><th>Var Comp.</th><th>% of Total</th></tr>
<tr><td>Batch</td><td>1756.799</td><td>86.25</td></tr>
<tr><td>Location</td><td>95.556</td><td>4.69</td></tr>
<tr><td>Rep</td><td>184.444</td><td>9.06</td></tr>
<tr><td>Total</td><td>2036.799</td><td></td></tr>
</table>
</td>
<td>1. To quantify how significantly random factors impact KPC's.
Random factors can include such things as repeatability of measurement instrument, location of sample, incoming lots, etc.
2. To get greater accuracy than provided by Snee plotting</td>
<td>1. Organize test in nested or hierarchical manner shown in top diagram (e.g. each batch is measured at three times and each time has two readings)
2. Use MINITAB'S Nested ANOVA to produce COV table
3. Compare percentages for each component. A very large value for one component may indicate a need to work to reduce its contribution to variability.</td>
</tr>
<tr>
<td>Cause & Effect Diagrams</td>
<td>


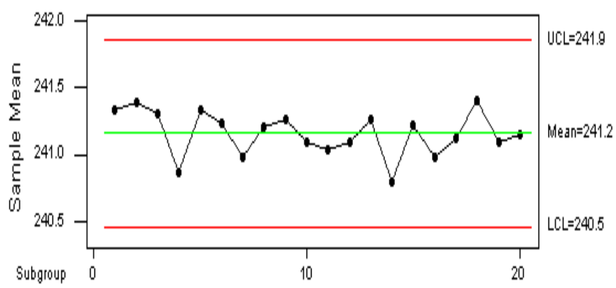
</td>
<td>1. Brainstorming tool for causes of certain effects (such as variability of process)
2. Narrow down causes</td>
<td>1. Draw fishbone diagram with effect visible along with categories of causes
2. Brainstorm with process experts on possible causes in each category
3. Select most probable causes to work on first</td>
</tr>
<tr>
<td>Best Fit Lines- Regression</td>
<td>


</td>
<td>1. You want to predict how a KPC is affected by another variable (e.g. Viscosity by Brine level)
2. You wish to use the relationship in a predictive way to find an optimum setting for a processing parameter</td>
<td>1. Vary processing parameter. Record KPC in production range
2. MiniTab's Stat Menu, choose Regression & then Fitted Line Plot...
3. Select closest line fit which is appropriate
4. Use equation to find optimum set points.</td>
</tr>
</table>

Statistical Tools – Handy Guide- Performance Qualification

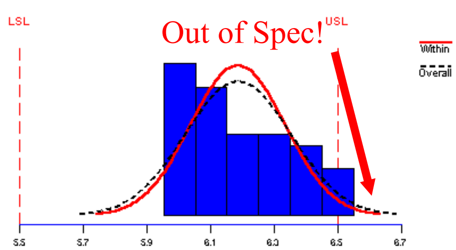
<table>
<tr><th>Tool</th><th>Description</th><th>Use When</th><th>Use How</th></tr>
<tr>
<td>Control Charts</td>
<td>Plot of variable over time compared to specific Control Limits

<table>
<tr><th>Name of Chart</th><th>Description</th></tr>
<tr><td>$\overline{X}$ Chart</td><td>Pronounced Xbar Chart. Groups of product are sampled and the Mean of the group plotted. For example 3 bottles are taken every hour and the Mean of the group recorded. This is often used in combination with an R chart.</td></tr>
</table>



Upper Control Limits (UCL), Central Line (CL) and Lower Control Limits (LCL)

UCL= $\overline{\overline{X}} + A_2\overline{R}$ CL= $\overline{\overline{X}}$ LCL= $\overline{\overline{X}} - A_2\overline{R}$

Where $\overline{\overline{X}}$ is the Mean of Means, $\overline{R}$ is mean of Range</td>
<td>1. Earlier PQ activities show process is stable & values are well within spec.
2. While setting or refining Control Limits as part of PQ
3. As part of regular production SPC to quickly locate & correct for:
a) Special Cause Variability
b) Increase in Common Cause Variability</td>
<td>1. Determine Control Chart type. See SPC Handy Guide (see top diagram)
2. Select Sample Size, Frequency, Number Samples
3. Enter your initial data in MiniTab*
4. Find Control Chart type in MiniTab's Stat Menu under Control Charts
5. Notice Control Limits calculated by MiniTab* (UCL & LCL middle, lower diagram)
6. Monitor process with control charts. Refine limits as process improves</td>
</tr>
<tr>
<td>Process Capability</td>
<td>

Potential (Within) Capability		Overall Capability	
Cp	1.12	Pp	1.00
CPU	0.70	PPU	0.62
CPL	1.54	PPL	1.37
Cpk	0.70	Ppk	0.62

PQ Process Capability Criteria

$C_{pk} < 1$
$1 \le C_{pk} \le 1.33$
$C_{pk} \ge 1.33$

Use after you demonstrate process "in-control" by Control Charts</td>
<td>1. Process is in-control according to Control Chart analysis
2. You believe that you will soon be ready to submit proof of process capability
3. You want to see not only if actual values (blue bars) fall within spec limits (LSL & USL) but also if expected values (normal curves) do
4. You notice or make a change to a capable process</td>
<td>1. Sample size ≥ 30, use MiniTab CAPA or Capability Six-pack
2. Check if any actual values (blue bars) or expected values (normal curves) fall outside spec limits
3. If process is in-control, use the values for Cp and Cpk to assess capability.
4. If process is not in-control, use the values for Pp and Ppk to assess capability along with Cpk – to – Ppk ratio.</td>
</tr>
</table>

* MINITAB ® is a registered trademark of MINITAB INC.