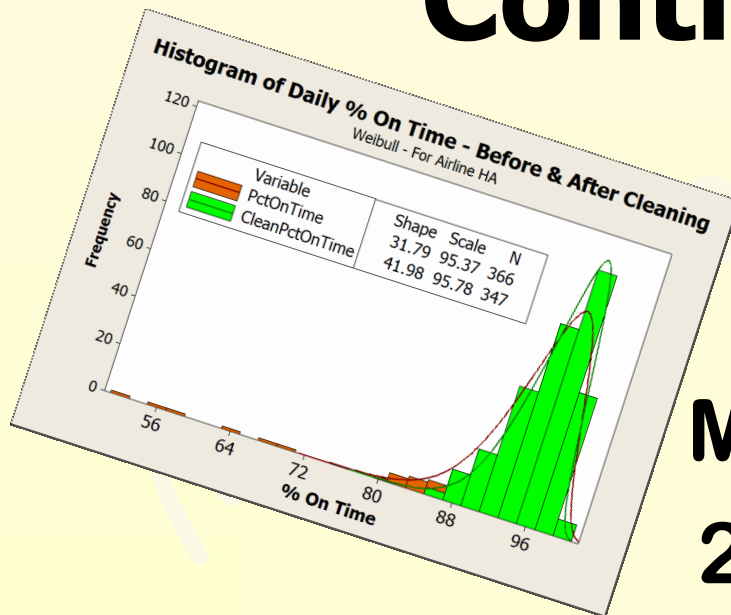


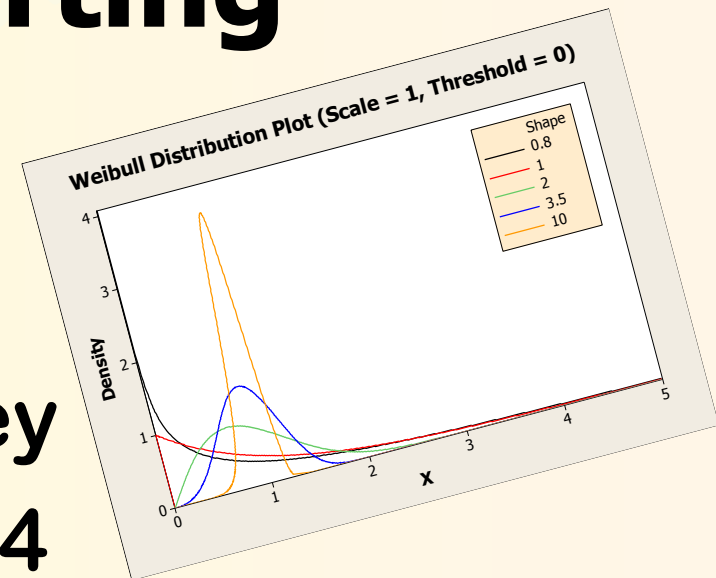


M.A.Sibley Consulting

Weibull Analysis as a Tool for Yield Analysis and Control Charting



Mark Sibley
2013-02-24



Weibull Analysis for Yield

The Weibull Distribution is the “go to” distribution for modelling **reliability data**.

It is also a distribution used to model the **volume output of a production process** where it can give you a sense of the “sprint capability” of the process and how much volume is being lost to special cause – essentially unreliability.

This training module will cover another use of the Weibull Distribution – that of fitting the distribution to **Yield** data.

Objectives of this module

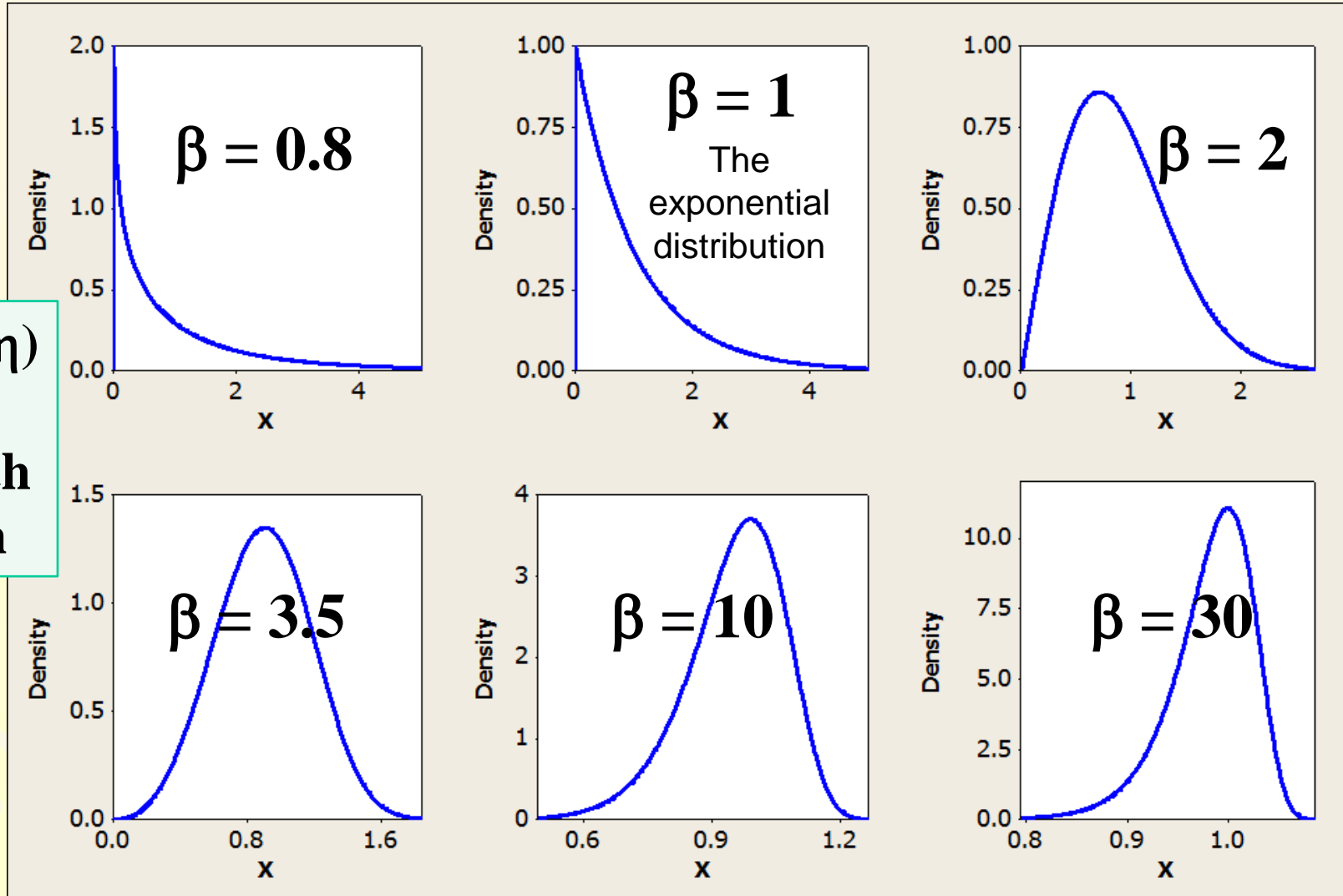
- To provide background on the **Weibull distribution**
- To outline how the Weibull distribution can be fitted to Yield data – e.g. the Shift or Daily Yield of a Manufacturing Process
- To outline a strategy for **automated determination of these parameters** and other relevant statistics
- To provide a practical approach to creating a custom control chart based on these parameters
- Discussion on how to **automate** creation of these **control charts**

The Weibull Distribution

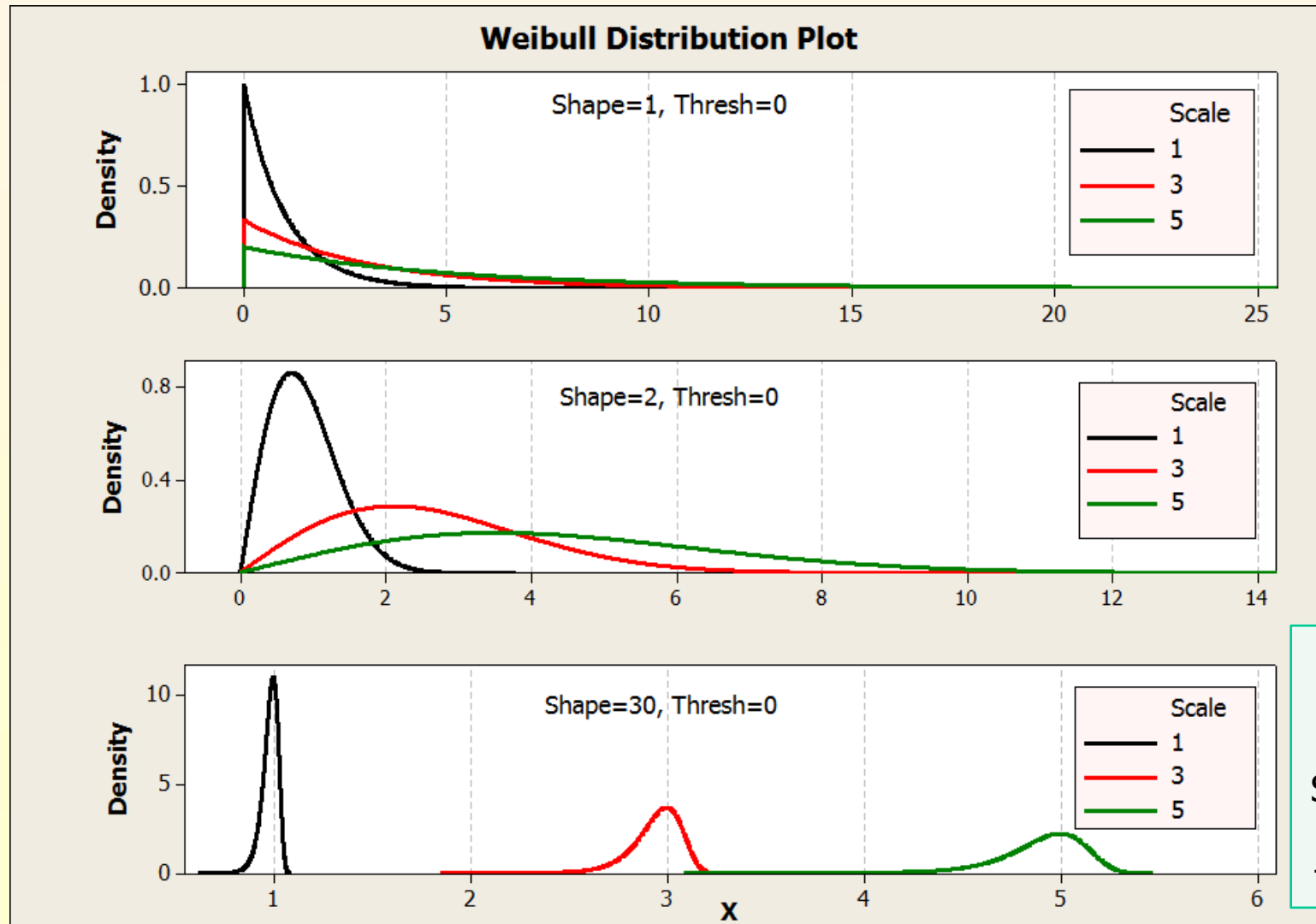
- The 2 Parameter Weibull Distribution is defined by:
 - Eta (η) or **Scale** – the point at which 63.2% of the area under the curve is to the left; for high Shape values this is *essentially the peak* of the distribution.
 - Beta (β) or **Shape** – changes the shape:
 - ≤ 1 is steadily decreasing
 - > 1 to 3 peaked with a tail to the right
 - 3 to 4 bell shaped
 - > 4 peaked with a tail to the left

Weibull Shape Parameter (β)

Scale (η)
= 1
for each
graph



Varying Scale and Shape



Note that at high Shape values, the Scale parameter is the peak of the distribution.

- It is helpful to think that yield data will include common cause and special cause variability.
 - **Common cause variability** will include the day to day sources of variability that will combine to create the “usual” variability which can often be well modeled by a Weibull Distribution
 - **Special cause variability** will include failures, start ups, shutdowns and other sources of variability which are unusual or are usual sources of variability but of an unusually large magnitude. If not dealt with, special cause variability will cause the overall yield data to be poorly fitted by a Weibull distribution.

- Trial and error methods can be used to try to fit a Weibull model to the underlying common cause data:
 - Trim the low end of the data before fitting to the distribution, review, repeat as needed. The assumption is that the low end represents “special cause” variability.
 - Adjust the parameters until the upper end of the distribution fits well to a Weibull Distribution.
- The problem is that a manufacturing plant may have dozens of machines or production lines producing hundreds of products. This can result in hundreds of machine – product combinations, each of which need to be modeled before a control chart based on a Weibull fit to the data can be created. Hand fitting is impractical.

Overview of Approach

- It is assumed here that you have manufacturing yield data organized by machines (or production lines). You can create Yield Control Charts based on the Weibull distribution:
 1. Lay out the groupings of machine vs machine centres vs products. (**Organize**).
 2. Use an **automated routine to calculate Weibull parameters** and the Xth Percentile of the Distribution (5th Percentile is a reasonable choice) for each *Machine Centre – Product* combination.

Overview of Approach – cont'd

3. Use those limits to **create or update a Weibull Limits file** that stores, at a minimum, the Scale and Xth Percentile value for each Machine Centre – Product combination.
4. **Create automated control charts** that read in data, read in limits and output control charts for each Machine – Product combination that is running.
5. **Schedule** the control charting routine to run each shift / day as appropriate and make a review of the chart part of the shift / day rituals.
6. **Periodically update** the control limits per step 2.

Step 1

Organize

1. Lay out the groupings of machine vs machine centres vs products. Simplified example:

Machine 1
(*Design Type "A"*)
Runs Products:
N23, N27, R5

Machine 2
(*Design Type "A"*)
Runs Products:
X1, N27

Machine 3
(*Design Type "A"*)
Runs Products:
X1, N27

Machine 4
(*Design Type "B"*)
Runs Products:
N27, Q17, R5

Machine 5
(*Design Type "B"*)
Runs Products:
N23, Q17

Assume that machines of a given design type should run products at equal yields – i.e. that they can be grouped together for Weibull fitting.

Based on this set up, the automated routine would calculate Weibull parameters for the following Machine Centre – Product Combinations:

A-N23, A-N27, A-R5, A-X1,
B-N27, B-Q17, B-R5, B-N23

Sample Data for this Module

Comments about the data used for this module:

- Ideally the data used would be manufacturing data but, not surprisingly, public data about the yield of a manufacturing process is hard to come by!
- At the site Data.GOV, you can get data on each scheduled airline flight in the U.S.
 - Data for each month in 2012 was extracted covering over 6 million flights
 - The data includes data on delays (Carrier, Weather, NAS, Security, Late Aircraft)
 - The data was summarized and “**On Time % Yield**” was defined for by each Carrier and Day as:

$$\frac{(All\ Flights - Delayed\ Flights)}{All\ Flights} * 100$$

← This is our “surrogate” for Manufacturing Process Yield.

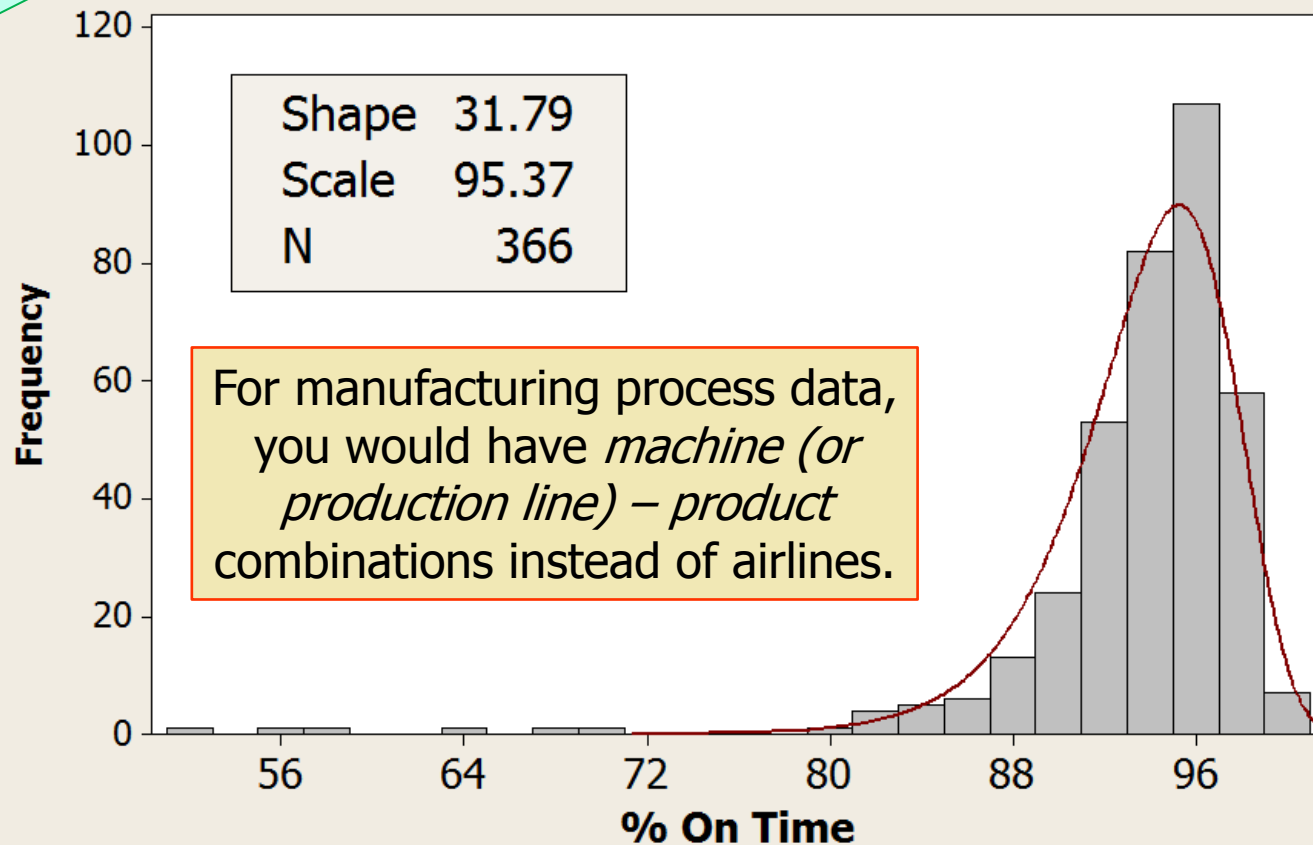
Step 2

Calculate Weibull Parameters

One Airline's Data

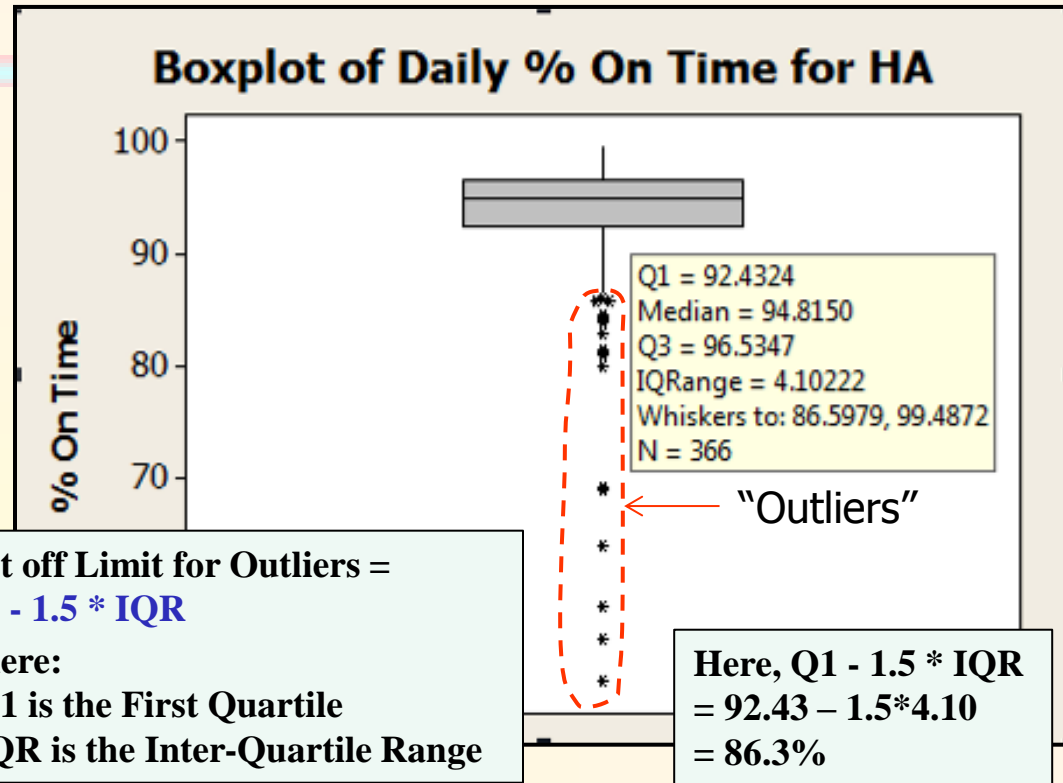
Histogram of Daily % On Time

Weibull - For Airline HA



The fit to a Weibull Distribution is not bad, but it looks like there is special cause data on the low side.

If we are to automate the estimation of the parameters for the Weibull fit, we need criteria for “Cleaning” the data i.e. to remove the “Special Cause” points and leave the “Common Cause” points. “Special Cause” here means the low end outliers. There are “shades of grey” in determining what to call an outlier, but one reasonable approach is that used in a BoxPlot. i.e. The criteria used here is to say anything that would be a low side outlier on a BoxPlot will be removed before estimating the Weibull parameters.

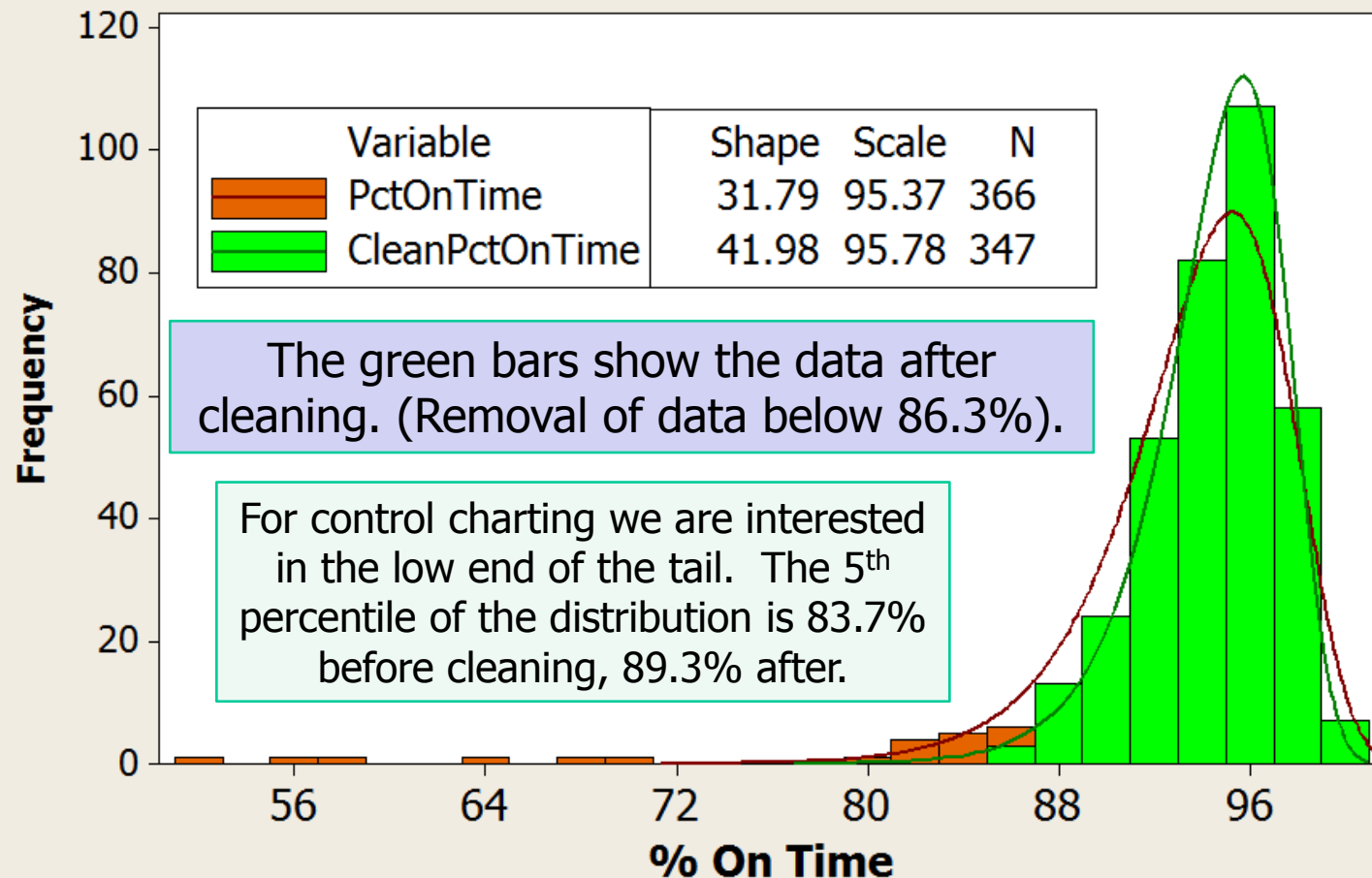




Before & After Cleaning

Histogram of Daily % On Time - Before & After Cleaning

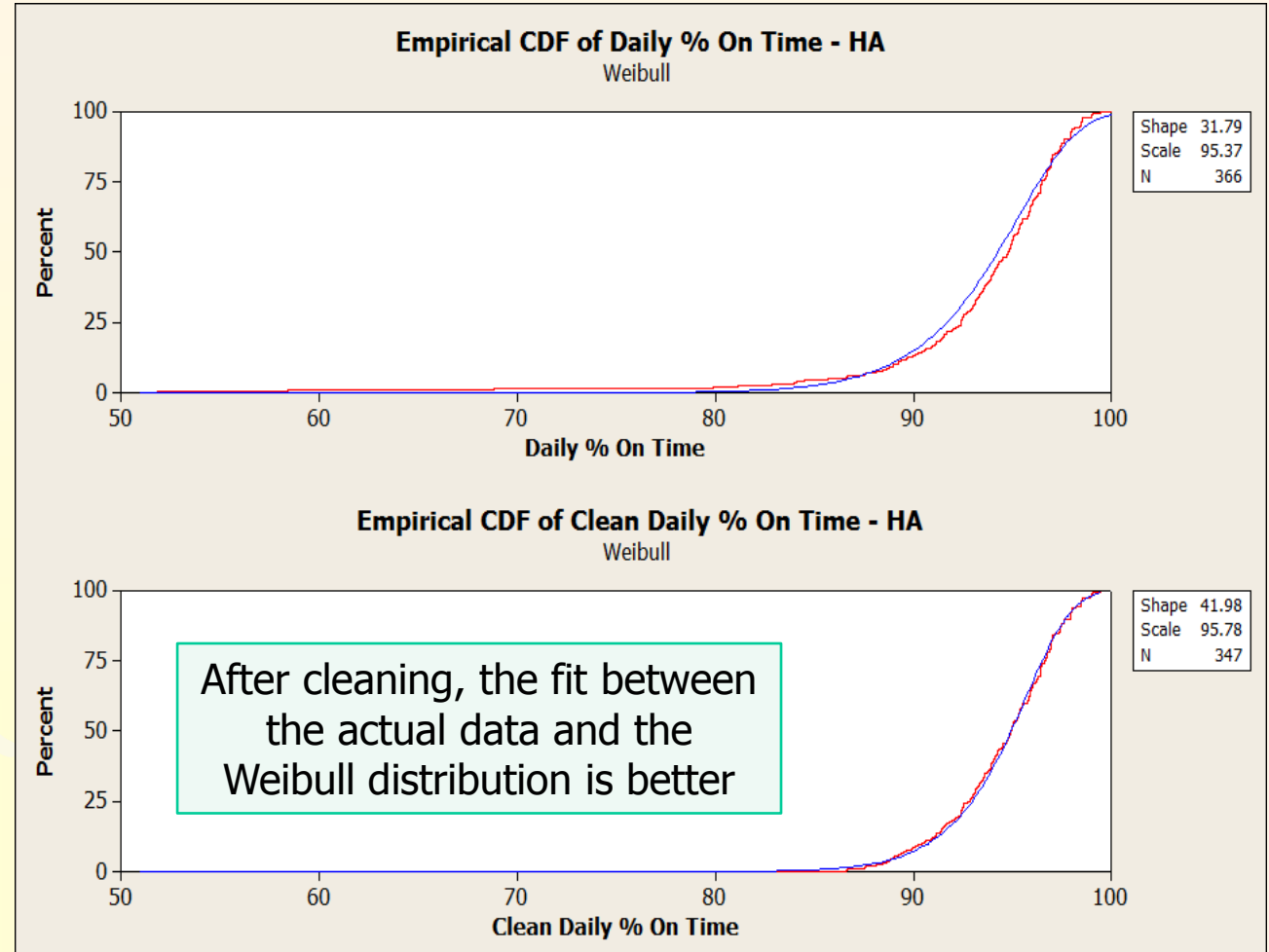
Weibull - For Airline HA



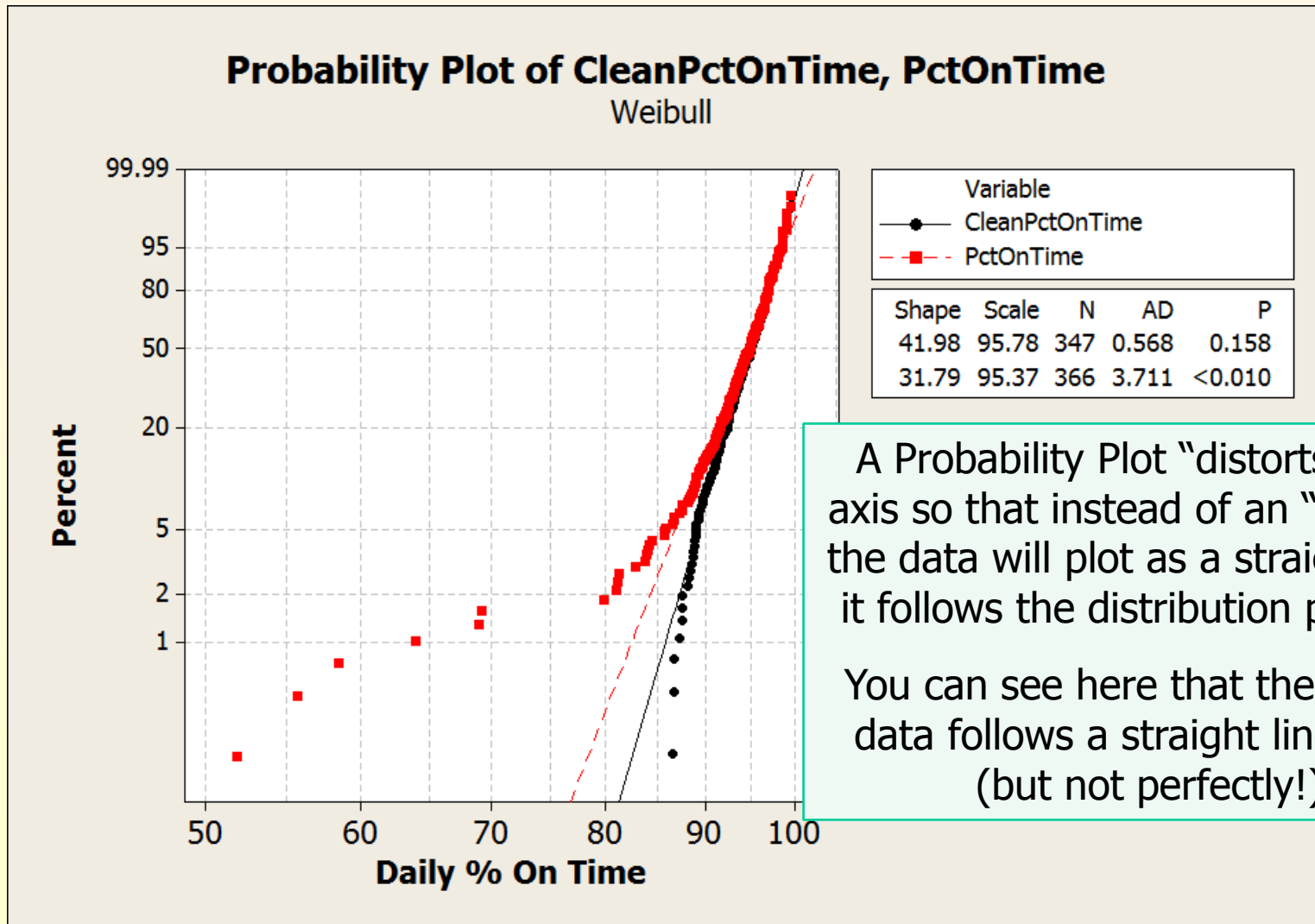
Cumulative Distribution

The cumulative distribution plot shows the actual % of data included up to each point on the X axis. At the left none of the data is included and by the right all of it is included.

The blue line shows the best fit Weibull distribution.



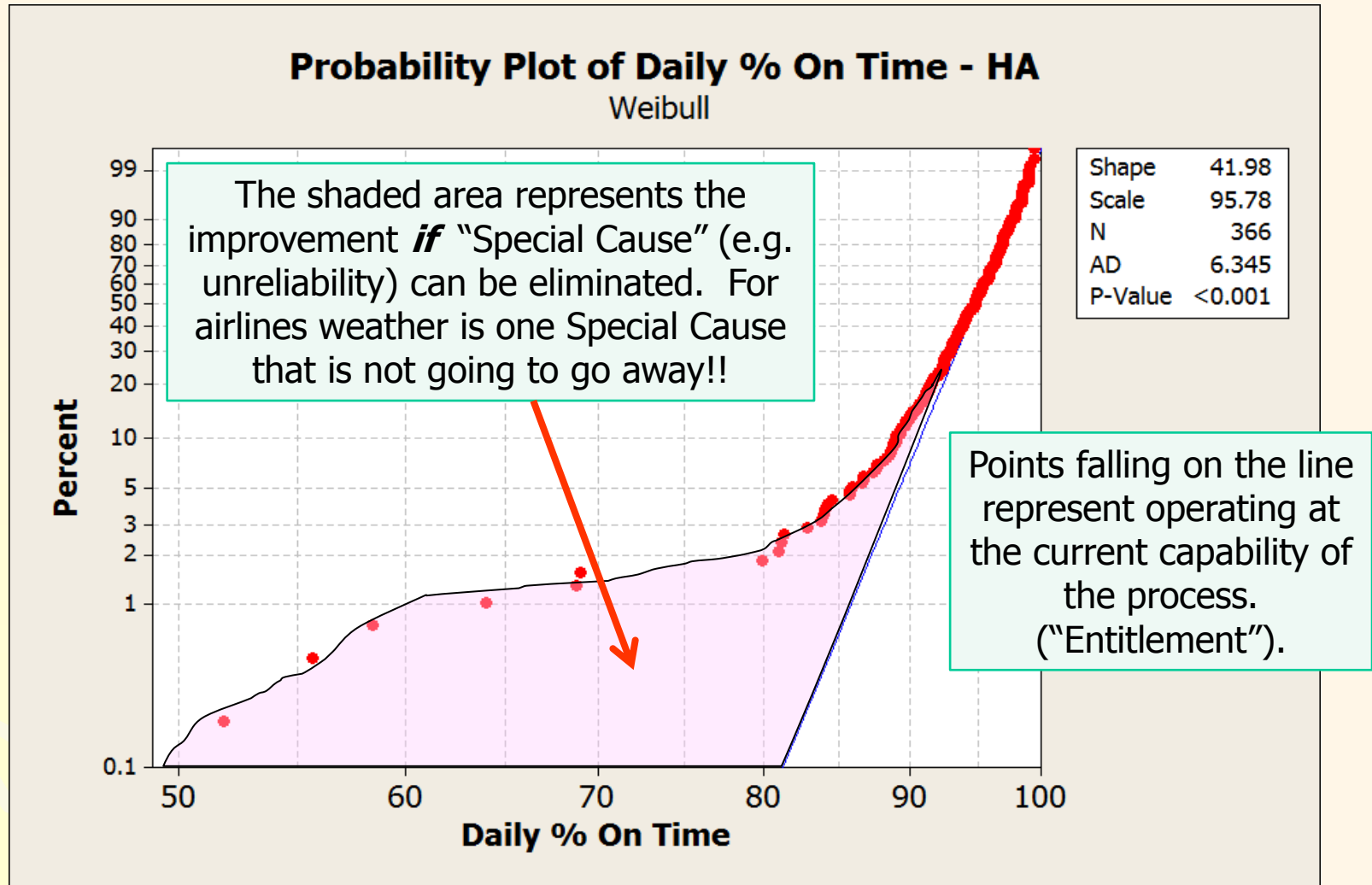
Probability Plot – Before & After



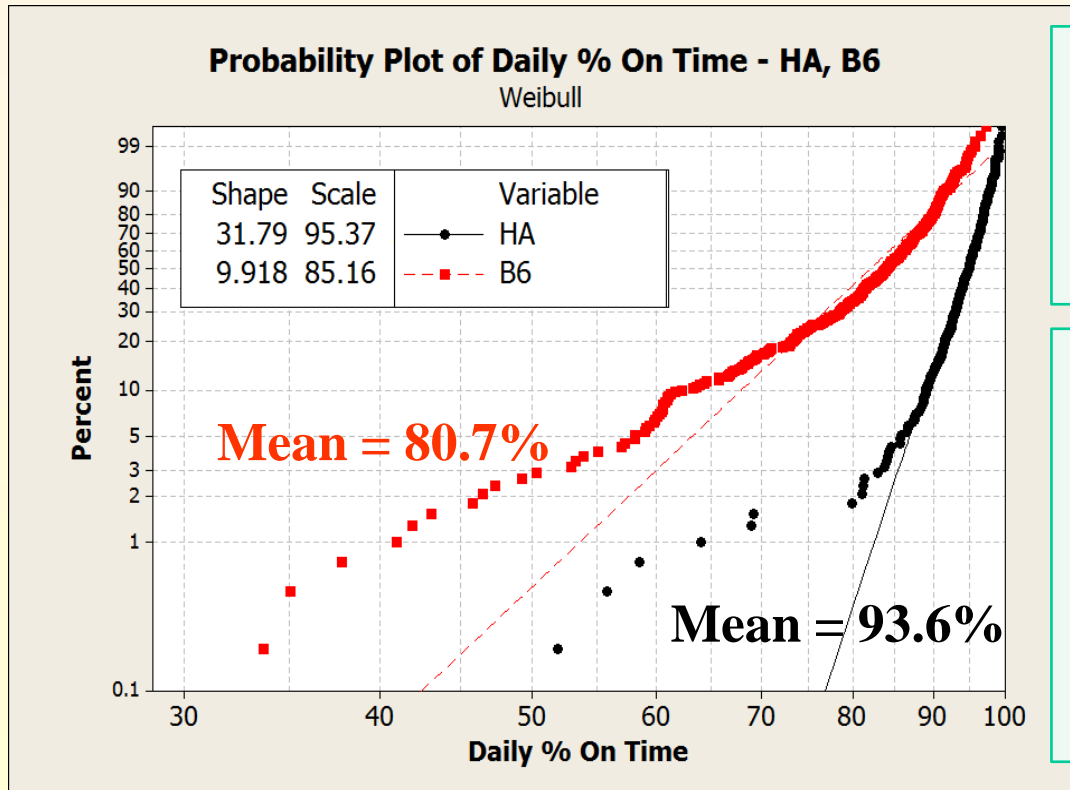
A Probability Plot “distorts” the Y axis so that instead of an “S” curve, the data will plot as a straight line if it follows the distribution perfectly.

You can see here that the cleaned data follows a straight line better (but not perfectly!).

Interpretation



Comparing Processes



These 2 different airlines have different capabilities. A higher Shape Parameter indicates the yield values are more crowded toward 100% - which is a good thing!!

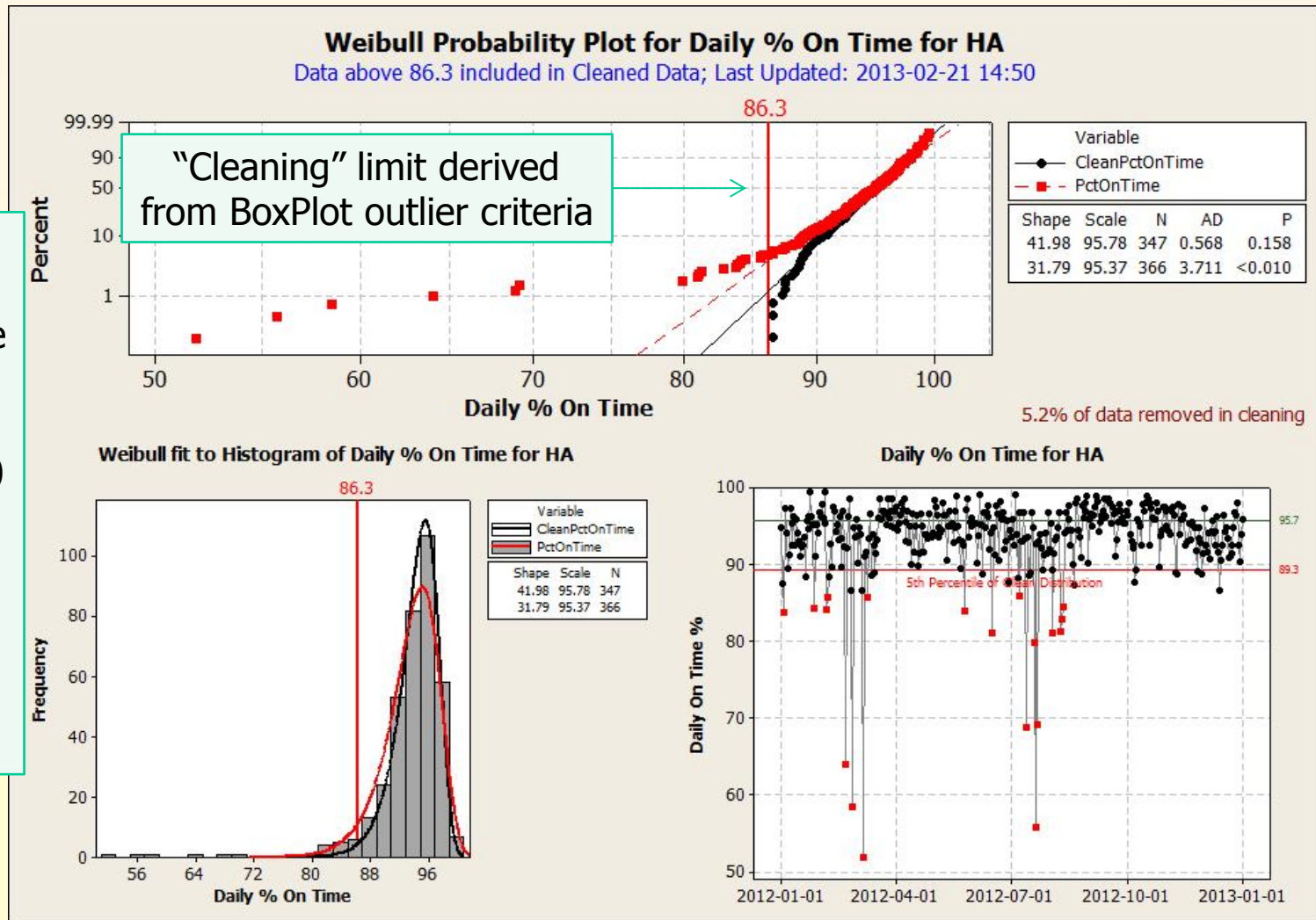
The Scale Parameter (for high values of the shape parameter) is essentially the peak of the distribution and can be considered the "Sprint" capability yield for the process i.e. when only common cause variability is in play yields should bounce around this value.

The mean yield is a function of slope of the line – the underlying capability of the process – **and** how much deviation there is from this line due to special cause variation. You can see in this example that the average % on time performances are very different. By making beneficial changes to the process, the underlying capability can be improved.

Step 2. Cont'd

Output of the Automated Routine

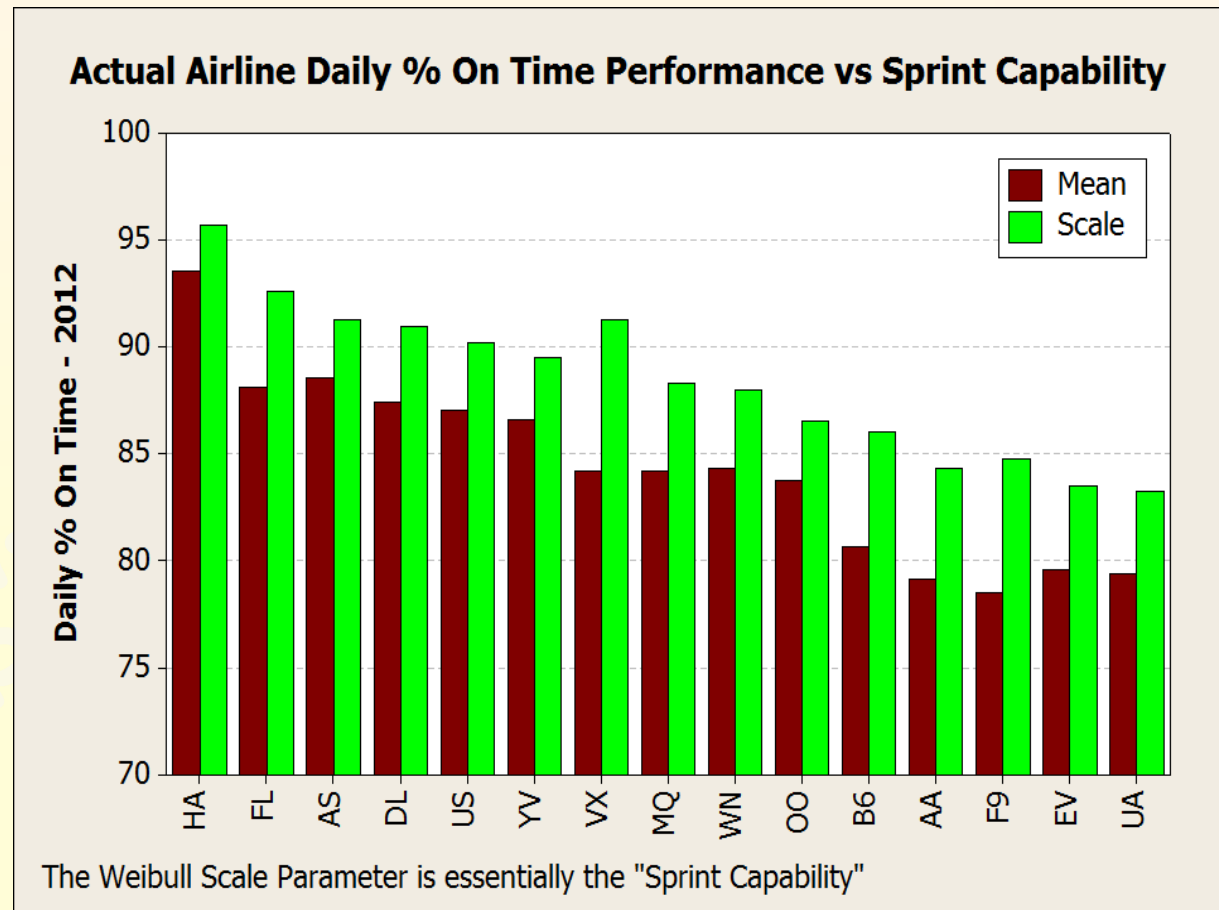
This graph comes from the automated routine which analyzes each combination (here each airline) to clean the data and capture the Shape and Scale parameters, 5th Percentile value, etc.



Overall “Yield” Summary

You are probably curious about how the airlines stacked up in 2012 so here is the summary graph. The scale parameters are for the cleaned data.

This graph is produced via the Automated Weibull Parameter Estimation macro.

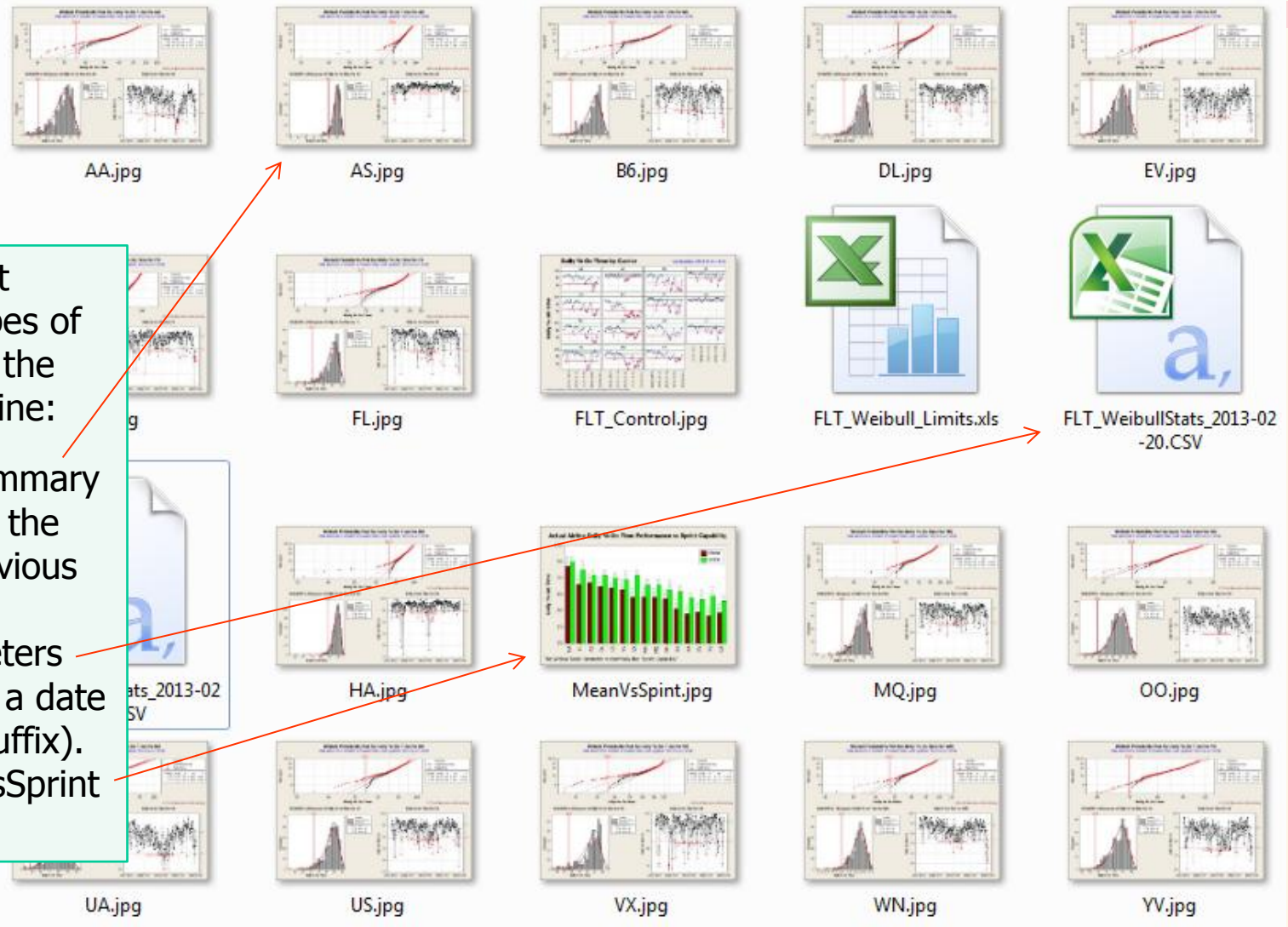


Step 2. Cont'd

Output of the Automated Routine

This screen shot shows the 3 types of files created by the automated routine:

1. The JPG summary graphs (like the one the previous slide).
2. The parameters file(s) (with a date of update suffix).
3. The MeanVsSprint graph.



Step 2. Cont'd

Output of the Automated Routine

- Contents of the CSV file containing the parameter estimates:

	A	B	C	D	E	F	G	H	
1	TCarrierNm	TScale	T5th	X	TPctCleaned	TMean	TMedian	TN	
2	AA	84.3	64.1		4.4	79.1	82.0	366	
3	AS	91.3	82.6		4.6	88.6	89.9	366	
4	B6	86.1	64.9		3.8	80.7	84.0	366	
5	DL	91.0	78.8		5.2	87.5	89.4	366	
6	EV	83.5	65.0		1.4	79.6	81.4	366	
7	F9	84.8	62.5		4.9	78.5	81.9	366	
8	FL	92.6	76.5		4.6	88.1	91.3	366	
9	HA	95.7	89.3		5.2	93.6	94.8	366	
10	MQ	88.3	70.9		2.7	84.2	86.4	366	
11	OO	86.6	74.8		2.7	83.8	84.4	366	
12	UA	83.3	64.4		1.1	79.4	80.9	366	
13	US	90.2	79.2		5.7	87.1	88.4	366	
14	VX	91.3	68.2		6.8	84.2	87.9	366	
15	WN	88.0	73.5		3.6	84.3	85.9	366	
16	YV	89.5	75.4		1.1	86.6	87.9	366	
17									

Step 3

Limits File

- Using the output from the Step 2. (automated parameter estimation), create or update the Limits file that will be used by the control charting routine.

	B	C	D	E	F	G	H	I
1	TScale	T5th	X	TPctCleaned	TMean	TMedian	TN	Comments
2	84.0	64.0		4.4	79.1	82.0	366	Updated 2013-02-23
3	91.0	82.5		4.6	88.6	89.9	366	Scale from 89.5 to 91.0 Updated 2012-12-15
4	86.0	65.0		3.8	80.7	84.0	366	
5	91.0	78.5		5.2	87.5	89.4	366	
6	83.5	65.0		1.4	79.6	81.4	366	
7	84.5	62.5		4.9	78.5	81.9	366	Updated 2013-02-23
8	92.5	76.5		4.6	88.1	91.3	366	
9	95.0	89.0		5.2	93.6	94.8	366	Analysis suggested Scale of 95.7, used 95
10	88.0	70.5		2.7	84.2	86.4	366	New Product - added 2012-10-13
11	86.5	74.5		2.7	83.8	84.4	366	
12	83.0	64.0		1.1	79.4	80.9	366	
13	90.0	79.0		5.7	87.1	88.4	366	
14	91.0	68.0		6.8	84.2	87.9	366	
15	88.0	73.5		3.6	84.3	85.9	366	
16	89.5	75.0		1.1	86.6	87.9	366	
17								

Step 4 Create the Control Chart

- Yield data are usually very non-normal
- A practical approach to alarming is:
 - Set a Lower Control Limit based on the Xth Percentile of the fitted Weibull Distribution – eg. 5th Percentile
 - Say that a point is in Alarm if:
 - The point is below the Lower Control Limit (LCL) *and*
 - The previous point is below the LCL

Step 4 Create the Control Chart

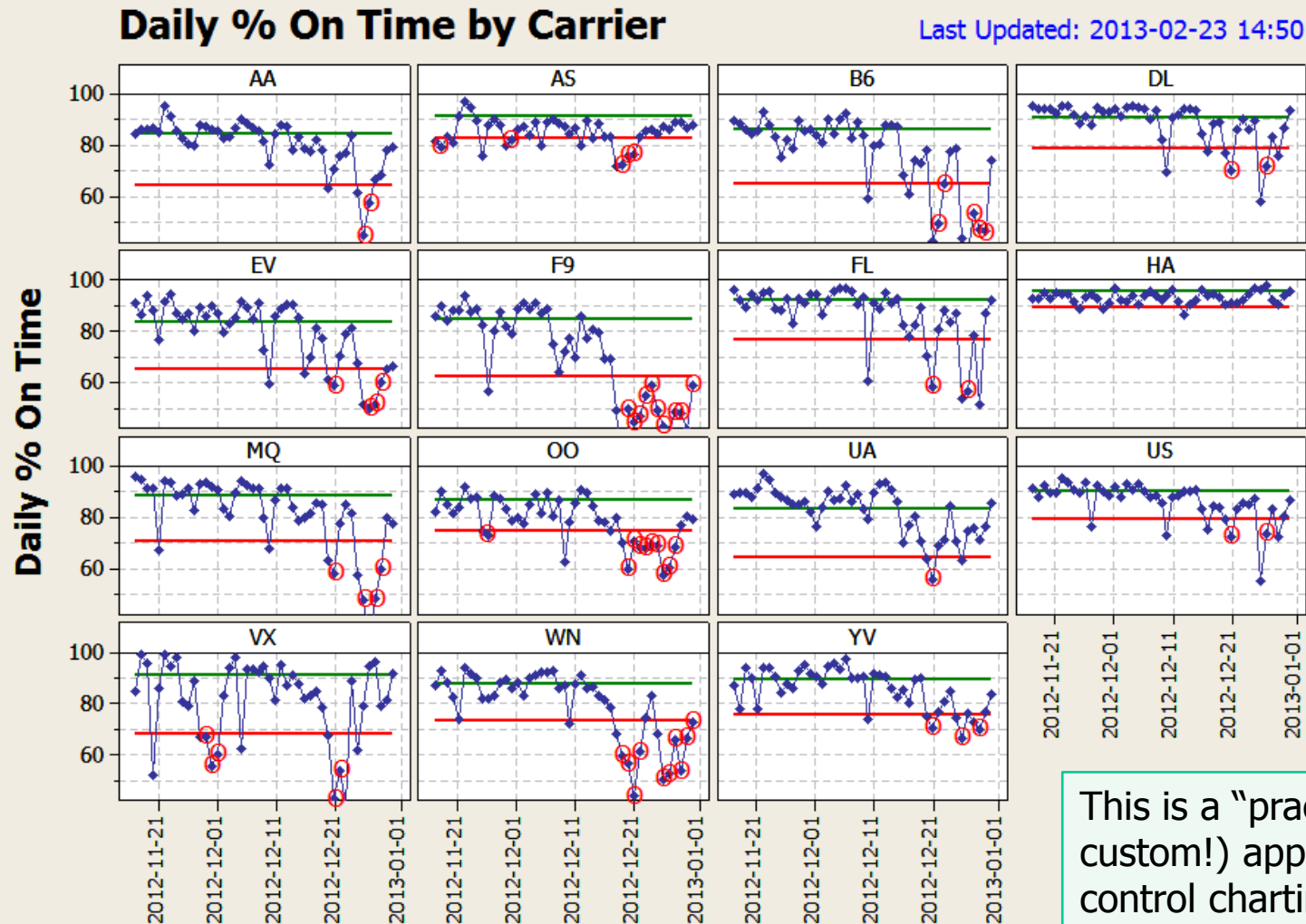
- If say 5% of the data had been cleaned and you use the 5th Percentile for the LCL, then 10% of the data should be below the LCL, making the chances of a single point being below the LCL to be 1 in 10. The chances that 2 points in a row (*if they were independent*) would be below the LCL would be about 1%. You can adjust the criteria to suit.
- This approach does not “ring the bell” for a single low yield data point. Often if you have very low yield for one day or shift, you know the reason or will start investigating right away.

Step 4 Create the Control Chart

- The macro to create the control chart should
 - Read in the latest data
 - Read in the limits file (with the Scale and Xth Percentile values).
 - Plot the latest results circling the points in alarm based on the criteria of 2 points in a row below the Lower Control Limit.
 - Store the control chart as a JPEG file to a shared drive so it can be linked to a web page and reviewed as needed.

Step 4.

The Control Chart created by Auto-Job



This is a "practical" (and custom!) approach to control charting.

- The web site <http://minitabmaestro.com/> has information on how Minitab “Auto-jobs” work.
- If you have a dynamic product wheel you may want to have the parameter estimation macro also run on a scheduled basis. That way if you see a new *machine centre – product* combination start running (i.e. there are no control limits on the control chart), then you can refer to the already analyzed parameter estimation data and use it to update the limits - as soon as you have enough data to have some confidence in the parameters.

Potential Roadblocks

- What if the automatic cleaning process does not yield a good fit to the Weibull Distribution?
 - You can try different criteria. However, if you have trouble “hand fitting” the data to a Weibull distribution, then you may need to look to a different distribution &/or approach
- Machines or Lines of same design vs different design – approach to estimating parameters
 - If you have multiple machines of the same design you should analyze them separately first to see if the assumption that the data can be lumped together and one set of parameters estimated for each machine – product combination is reasonable. Parameters based on *Machine Design Type – Product* combinations have the advantage that if a product moves to a new machine of the same design type then you can start control charting right away.

The Macro Code – Parameter Estimation Routine

```
# FLT_Weibull.mac - Process Summarized US Flight data creating Weibull Stats
#
#           M.A. Sibley Consulting, Kingston, Ontario.  Copyright 2013
#
# 2013-02-18 MAS - New Macro

gmacro
▢FLT_Weibull

    echo                # Helpful for debugging

    Call ReadInALL      # Summary data found in 12 CSV files
    Call ProcesByCarrier # Loop through carriers creating table & summary graphs
    Call WriteStats     # Write out table of Weibull stats vs Airline as CSV
    Call GraphMeanVsSprint # Graph of overall means vs scale parameters

endmacro
```

My text editor (EditPad Pro) provides the colour coding – which is very helpful for catching typos and improves the readability of the code immensely. The code shots that follow **ONLY show the comments**. The detailed code has been “folded”. (Another feature of the text editor).

The Macro Code – Parameter Estimation Routine

```
#####  
# ReadInALL  
#####  
  
gmacro  
ReadInALL  
  
# Read in previously consolidated files - consolidated file created via DOS  
# command: Copy S_2012*.csv T_2012.csv  
  
# DOS copy command will include variable names with each file concatenated so  
# show up as rows within the dataset. Get rid of them.  
  
# Close no longer needed worksheet  
  
endmacro
```


The Macro Code – Parameter Estimation Routine

```
#####
#  ProcesByCarrier
#####

gmacro
]ProcesByCarrier

    # Put Date in International Date Format

    # Create table of Carriers - this will be used in DO loop to process each
    #   airline

    # Round the output

    # Define number of carriers (Number of times to Loop)

    # Create the sub-title

    # Loop Through Processing 1 carrier at at time

    DO k31 = 1 : k30

        # Add Results into table of original data

        # Define stats for this carrier (table will be 1 row long)

        # Threshold Limit for cleaning - same as criteria for BoxPlot i.e.
        #   First Quartile minus 1.5 * InterQuartileRange

        # Create sub-title text

        # Create "Cleaned" version of variable - i.e. with outliers removed

        # Create status flag for variable

        # Calculate stats for cleaned variable - just the count
```

The Macro Code – Parameter Estimation Routine

```

# Create footnote

# Create filename for graph

# Create graph layout with 3 sub graphs; store JPG of graph to disk

layout;
  gsave k36;
  replace;
  jpeg.

# Use Stat > Reliability/Survival > Distribution (Right Censoring) > &
#   Parametric Distribution Analysis...
# to define the Weibull stats for the original and cleaned data
# (This analysis can create a graph but Graph > Probability Plot produces
#   a similar output but lets you control its location in a layout, etc.)

# Create a Probability Distribution Plot of the original and cleaned data
# (This is created interactively via Graph > Probability Plot)

# Create scatterplot of On Time performance vs Day

# Create a Histogram of the original (with bars) and cleaned data (no bars)

EndLayout

# Store the stats for the carrier along side the previously
#   created table of carriers

Call StoreStats

ENDDO

endmacro

```

The Macro Code – Parameter Estimation Routine

```
#####
# StoreStats
#####

gmacro
|StoreStats

|   # Define columns
|
|   # Populate Columns - one row each loop iteration - round unneeded decimals
|
|   endmacro
#####
# WriteStats
#####

gmacro
|WriteStats

|   # Use sort to move into first columns of a new worksheet
|
|   # Two columns have the mean (TOnTimeYield, TMean)
|   # Overwrite the duplicate information column to create a gap in the table
|   # The column with missing values needs to named otherwise it will not output
|   #   to CSV file
|
|   # Create the Filename Suffix
|
|   # Change into directory so file name will not be too long
|   #   This is one of the very few Minitab commands that Lets you interact
|   #   directly with the operating system
|
|   # Save worksheet as a CSV file
|
|   endmacro
```

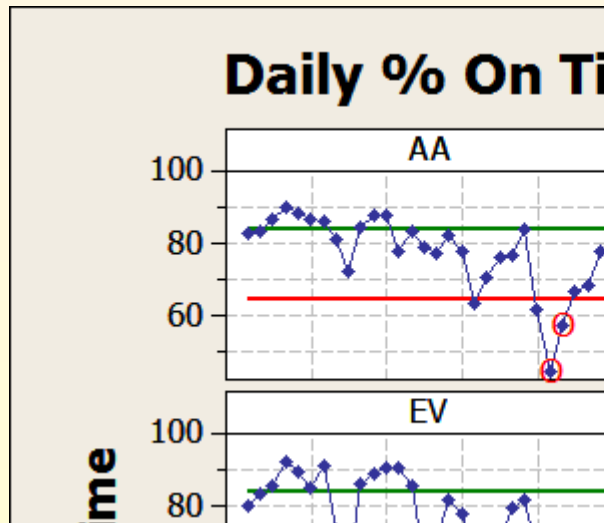
The Macro Code – Parameter Estimation Routine

```
#####  
# GraphMeanVsSprint  
#####  
  
gmacro  
GraphMeanVsSprint  
  
# Create bar chart showing actual mean and "sprint capability" by carrier  
  
endmacro
```



M.A.Sibley Consulting

The Macro Code – Control Charting



```
# FLT_Weibull.mac - Create example Weibull Control charts from US Flight data
#
# M.A. Sibley Consulting, Kingston, Ontario. Copyright 2013
#
# 2013-02-23 MAS - New Macro
```

```
gmacro
FLT_Weibull
```

```
echo # helpful for debugging
```

```
Call ReadInALL
```

```
Let c16[1] = '*' # so ODBC command will return columns to the right
```

```
Call ReadLimits # will return table into same worksheet
```

```
Call ControlChart # generate control chart from latest data
```

```
# Close no longer needed worksheet
```

```
endmacro
```

```
#####
# ReadInALL
#####
```

```
gmacro
ReadInALL
```

```
# Read in consolidated file - created via DOS command Copy S_2012*.csv T_2012.csv
```

```
# DOS copy command will include variable names with each file concatenated.
```

```
# Get rid of them.
```

```
# Close no longer needed worksheet
```

```
endmacro
```



The Macro Code – Control Charting cont'd

```
#####  
# ReadLimits  
#####  
  
gmacro  
ReadLimits  
  
    # Read in table via ODBC Command  
  
endmacro  
  
#####  
# ControlChart  
#####  
  
gmacro  
ControlChart  
  
    # Put Date in International Date Format  
  
    # Sort in place to get a given carrier's flights on successive rows  
  
    # Define if a point is in alarm; store the value "O" - upper case "Oh" to be  
    # used as a data label. (If it is centred over the point it looks like a  
    # circular plotting symbol).  
  
    # Create the sub-title  
  
    # Create control chart via layout; top layer is just to get points in alarm  
    #   circled  
  
    Layout;  
        gsave "C:\MASibley Consulting\Training\Reference\US Flight Data\Weibull\FLT_Control.jpg";  
        replace;  
        jpeg.  
  
    # Plot bottom layer - control chart less circled points  
  
    EndLayout  
  
endmacro
```