

StatSoft®

STATYSTICA

Analiza danych • Data mining • Sterowanie jakością • Analityka przez Internet

PROCEED - modelowanie, optymalizacja i symulacja złożonych procesów produkcyjnych

**Michał Iwaniec
StatSoft Polska**

„Zastosowania analizy danych w przemyśle”

Warszawa 2005

PROCEED

- ❑ Gotowe rozwiązanie dla zagadnień produkcyjnych
- ❑ Wyodrębnienie zależności pomiędzy produktem a procesem jego wytwarzania
- ❑ Implementacja opatentowanego podejścia, którego skuteczność została wielokrotnie potwierdzona

PROCEED to przepis

- ❑ Wypracowany sposób postępowania
- ❑ Użytkownik porusza się w sprawdzonym i wydajnym schemacie
- ❑ Six Sigma a dane

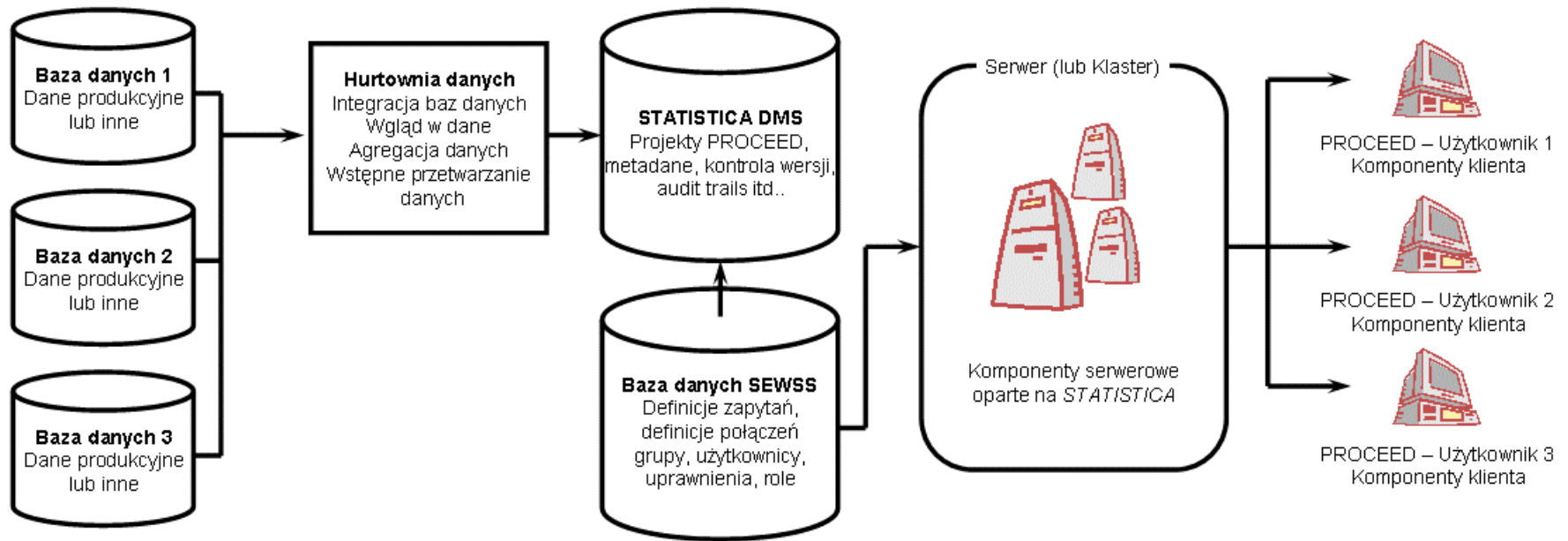
PROCEED - zastosowania

- ❑ Skomplikowane, wieloetapowe procesy produkcyjne
- ❑ Problemy podczas tworzenia modeli
- ❑ Produkt powinien spełniać kilka przeciwstawnych żądań
- ❑ Brak możliwości zastosowania tradycyjnych metod analitycznych

PROCEED i Six Sigma

- ❑ Usystematyzowanie przepływu informacji
- ❑ Wydobywanie, gromadzenie, rozpowszechnianie wiedzy
- ❑ Wysoka jakość samego procesu projektowania i wdrażania projektów
- ❑ Bezpieczeństwo

Architektura systemu



„Po pierwsze nie szkodzić”

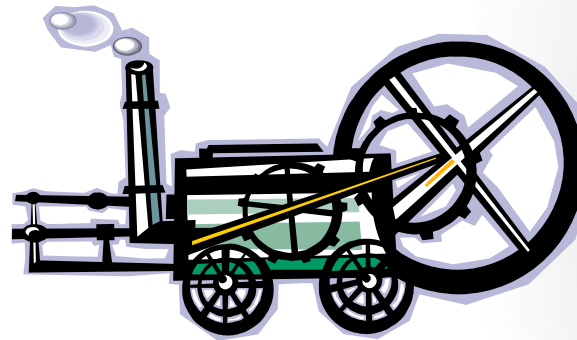
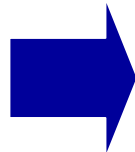
- Dobrze zdefiniowana i sprawdzona strategia dla:
 - Budowy modeli rzeczywistych procesów
 - Przekazywania informacji do odpowiednich osób
 - Monitorowania jakości stworzonych wcześniej modeli
 - Zarządzania modelami i innymi dokumentami

Proces budowy modelu

- W PROCEED budowa modelu składa się z następujących kroków:
 - Definiowanie danych
 - Czyszczenie danych
 - Usuwanie redundancji zmiennych
 - Redukcja wymiarów
 - Budowa modelu
 - Symulacja
 - Optymalizacja
 - Wdrożenie

Zdefiniowanie problemu

ok. 100 zmiennych
wejściowych

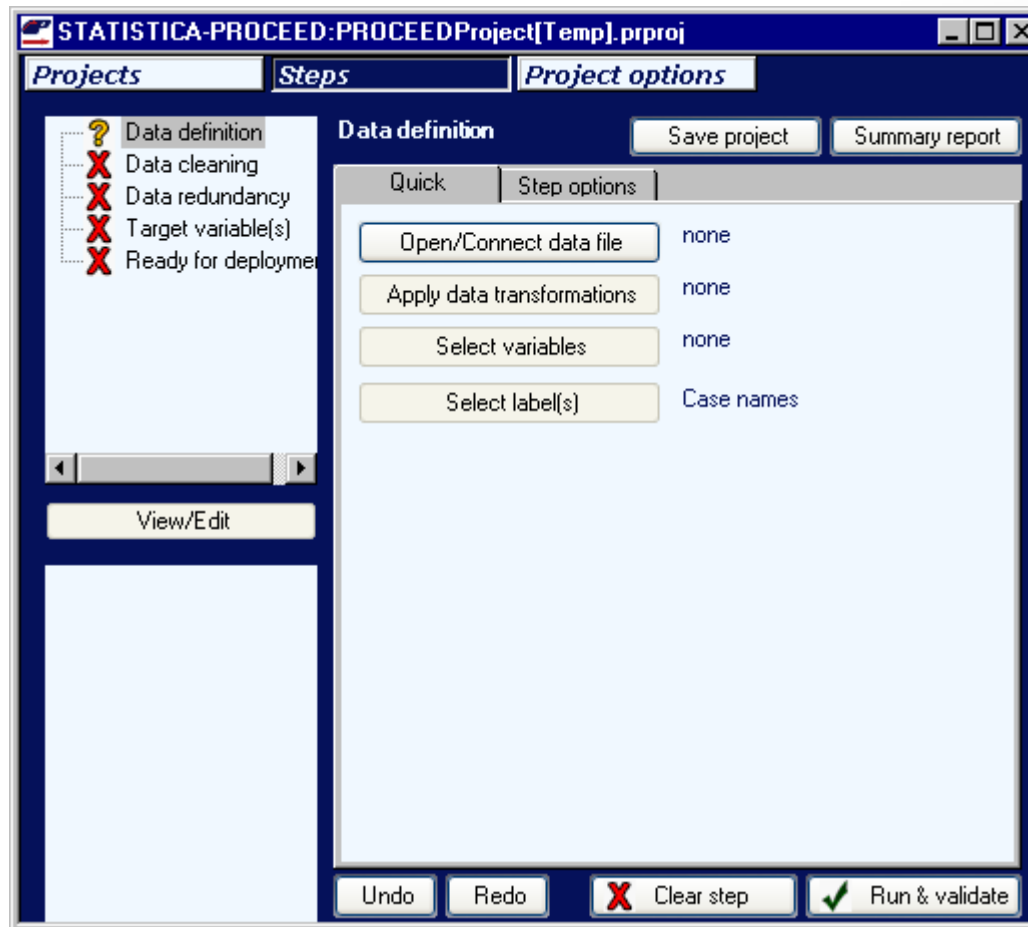


Sprawność

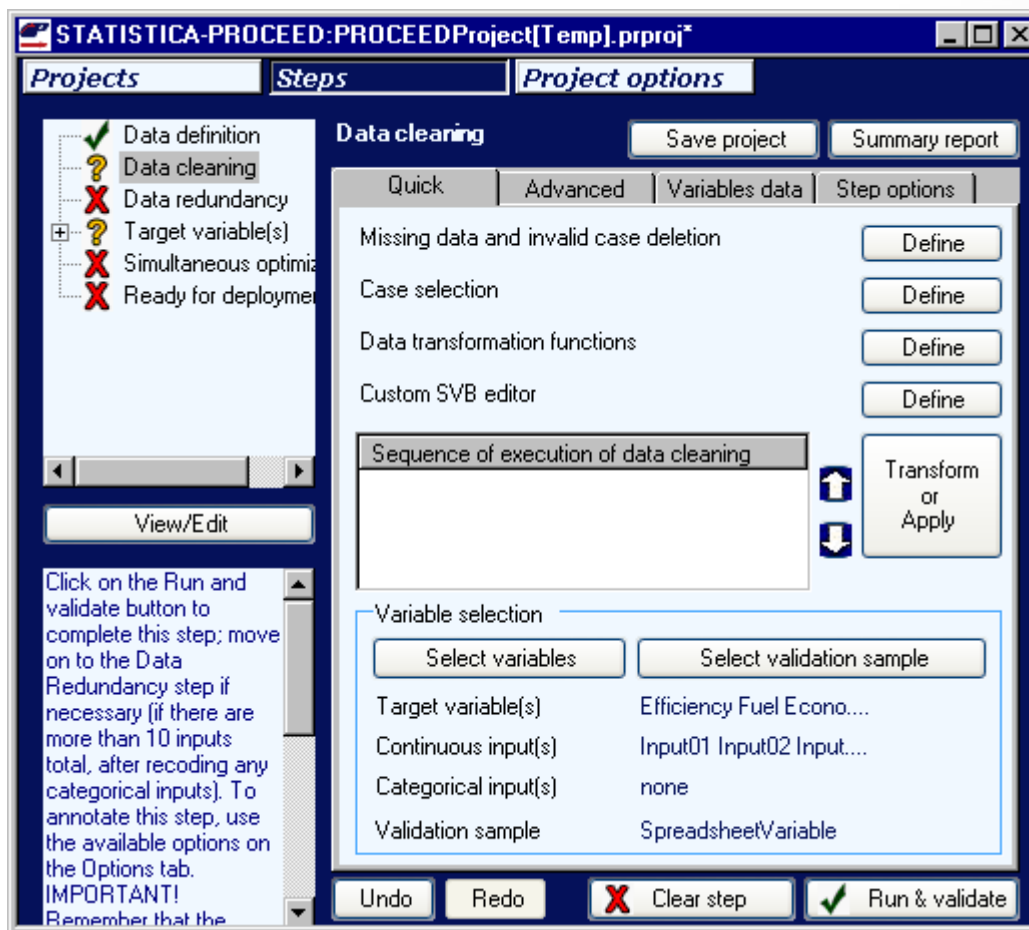
Moc

Zużycie
paliwa

Definiowanie danych



Czyszczenie danych

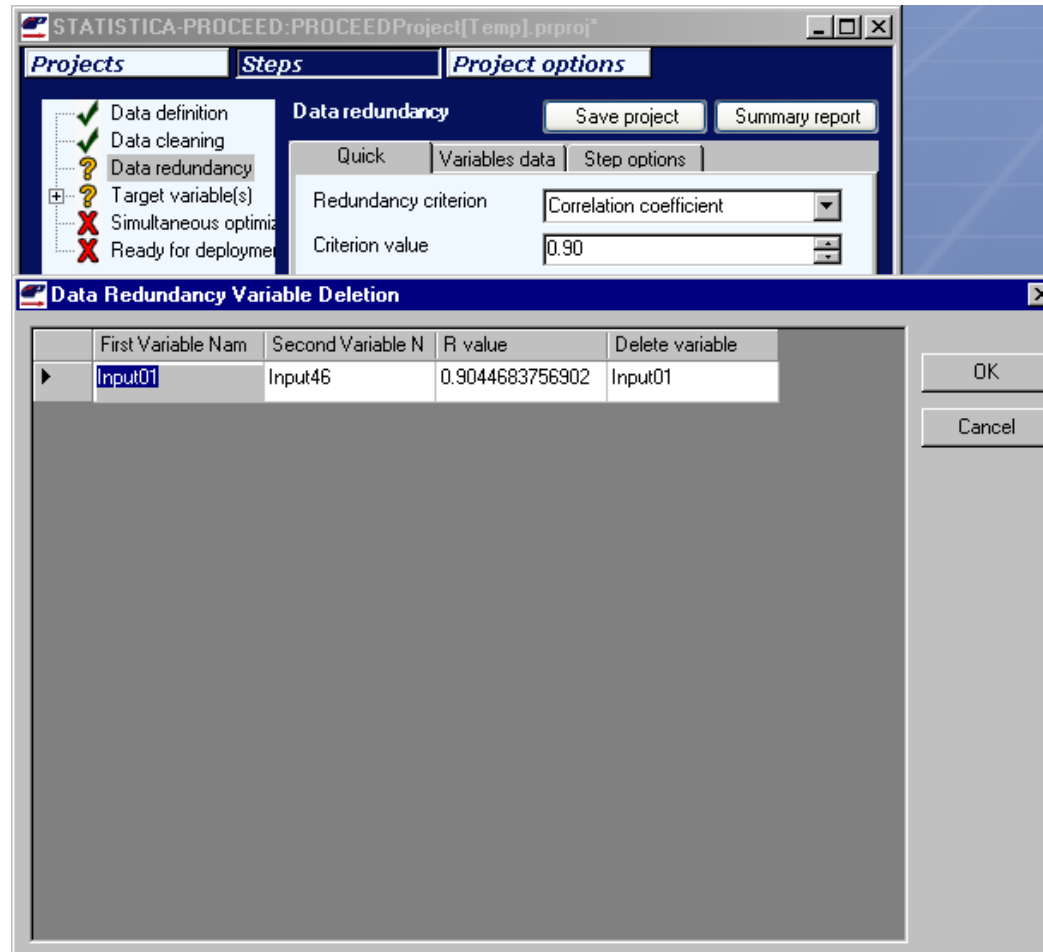


The screenshot shows the 'STATISTICA-PROCEED:PROCEEDProject[Temp].prproj*' window with the 'Steps' tab selected. The 'Data cleaning' step is highlighted in the 'Steps' list on the left. The main area displays the configuration for this step, including options for 'Missing data and invalid case deletion', 'Case selection', 'Data transformation functions', and 'Custom SVB editor'. The 'Sequence of execution of data cleaning' section is currently empty. The 'Variable selection' section shows 'Select variables' and 'Select validation sample' buttons, with the following settings:

Target variable(s)	Efficiency Fuel Econo...
Continuous input(s)	Input01 Input02 Input...
Categorical input(s)	none
Validation sample	SpreadsheetVariable

At the bottom of the window, there are buttons for 'Undo', 'Redo', 'Clear step' (with a red X icon), and 'Run & validate' (with a green checkmark icon). A 'View/Edit' button is located below the 'Steps' list. A help text box at the bottom left provides instructions on how to complete the step and use the 'Options' tab.

Usuwanie redundancji zmiennych



STATISTICA-PROCEED:PROCEEDProject[Temp]_prproj

Projects | **Steps** | **Project options**

- ✓ Data definition
- ✓ Data cleaning
- ? Data redundancy
- + ? Target variable(s)
- ✗ Simultaneous optimization
- ✗ Ready for deployment

Data redundancy | Save project | Summary report

Quick | Variables data | Step options

Redundancy criterion: Correlation coefficient

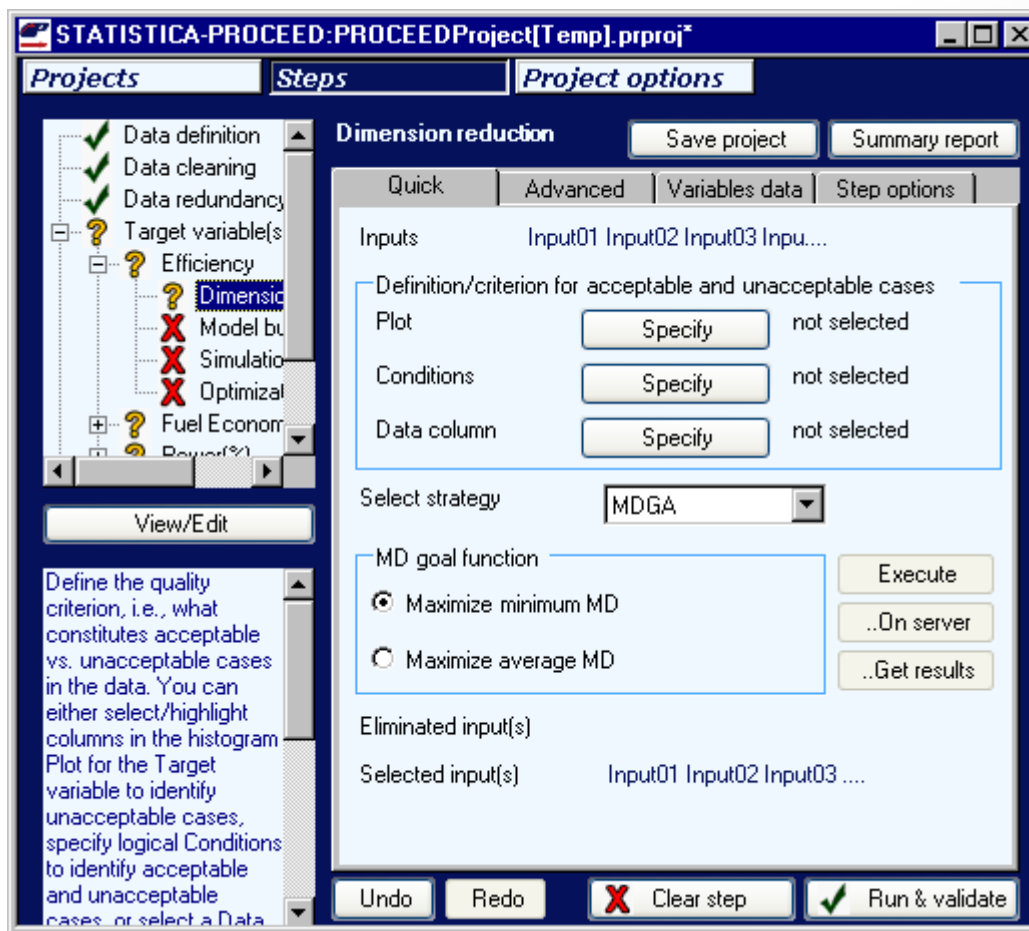
Criterion value: 0.90

Data Redundancy Variable Deletion

First Variable Name	Second Variable Name	R value	Delete variable
Input01	Input46	0.9044683756902	Input01

OK
Cancel

Redukcja wymiarów



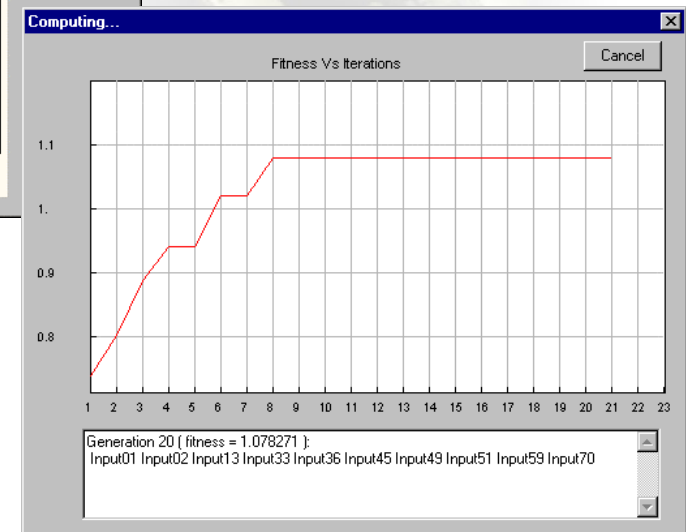
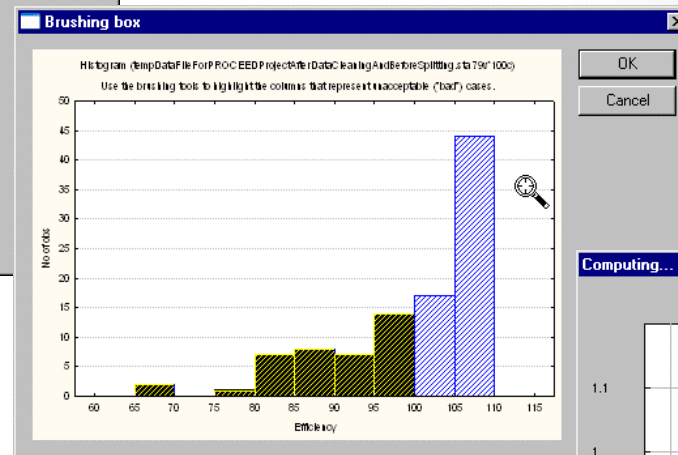
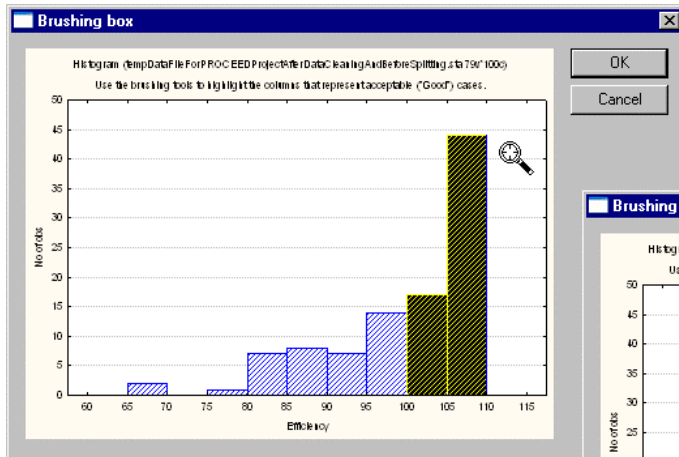
The screenshot shows the 'STATISTICA-PROCEED:PROCEEDProject[Temp].prproj*' window. The 'Steps' tab is active, and the 'Dimension reduction' step is selected in the 'Project options' pane. The 'Dimension reduction' dialog box is open, showing the 'Quick' tab. The 'Inputs' section lists 'Input01', 'Input02', and 'Input03'. The 'Definition/criterion for acceptable and unacceptable cases' section includes 'Plot', 'Conditions', and 'Data column', each with a 'Specify' button and 'not selected' status. The 'Select strategy' dropdown is set to 'MDGA'. The 'MD goal function' section has two radio buttons: 'Maximize minimum MD' (selected) and 'Maximize average MD'. The 'Eliminated input(s)' section is empty, and the 'Selected input(s)' section lists 'Input01', 'Input02', and 'Input03'. The 'View/Edit' button is visible in the 'Steps' pane. The bottom of the dialog box has 'Undo', 'Redo', 'Clear step', and 'Run & validate' buttons.

Algorytm genetyczny

- ❑ Losowanie populacji początkowej
- ❑ Ocena populacji
- ❑ Operatory ewolucyjne
 - krzyżowanie
 - mutacja
- ❑ Powstaje nowa populacja

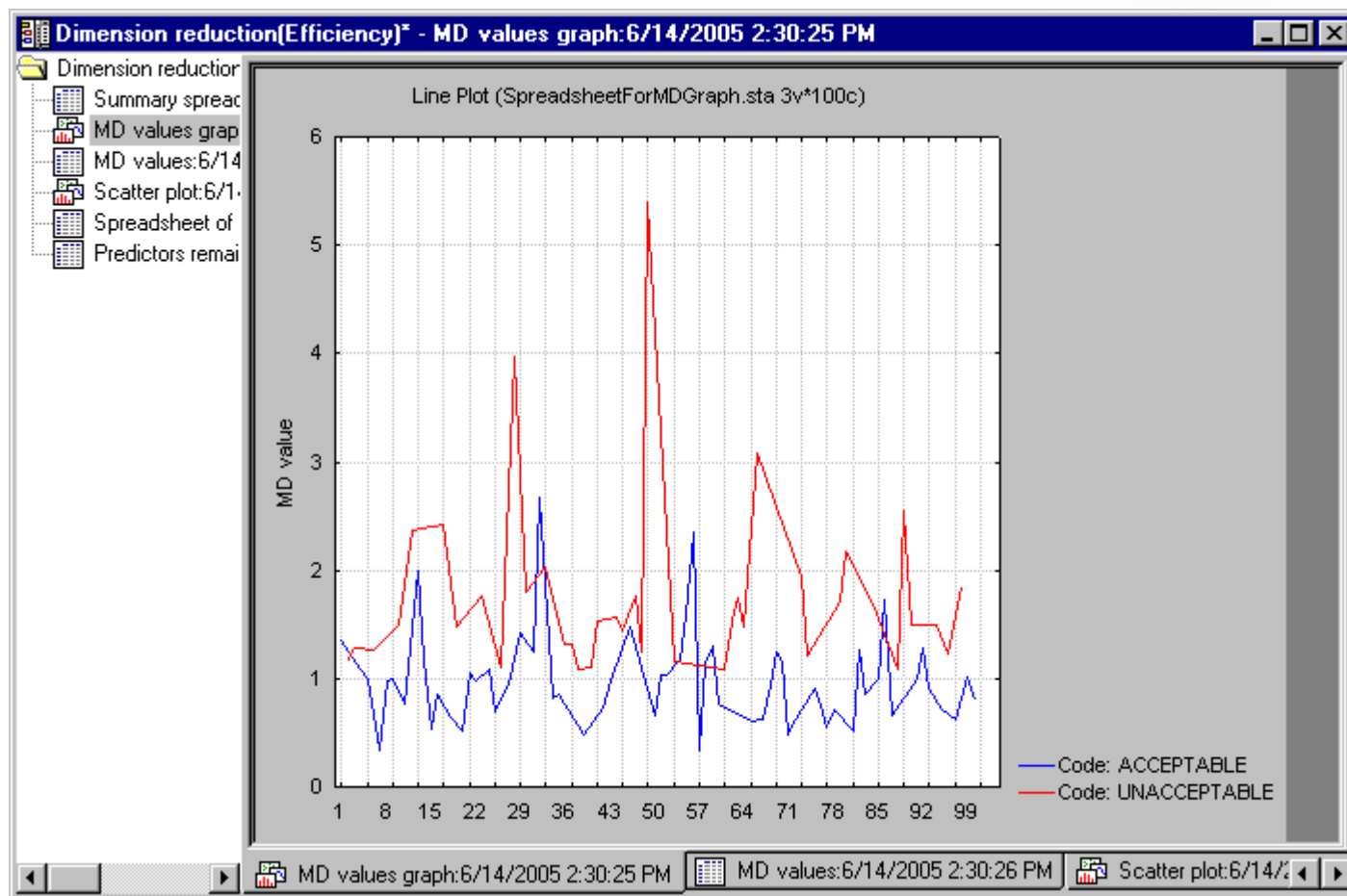


Redukcja wymiarów

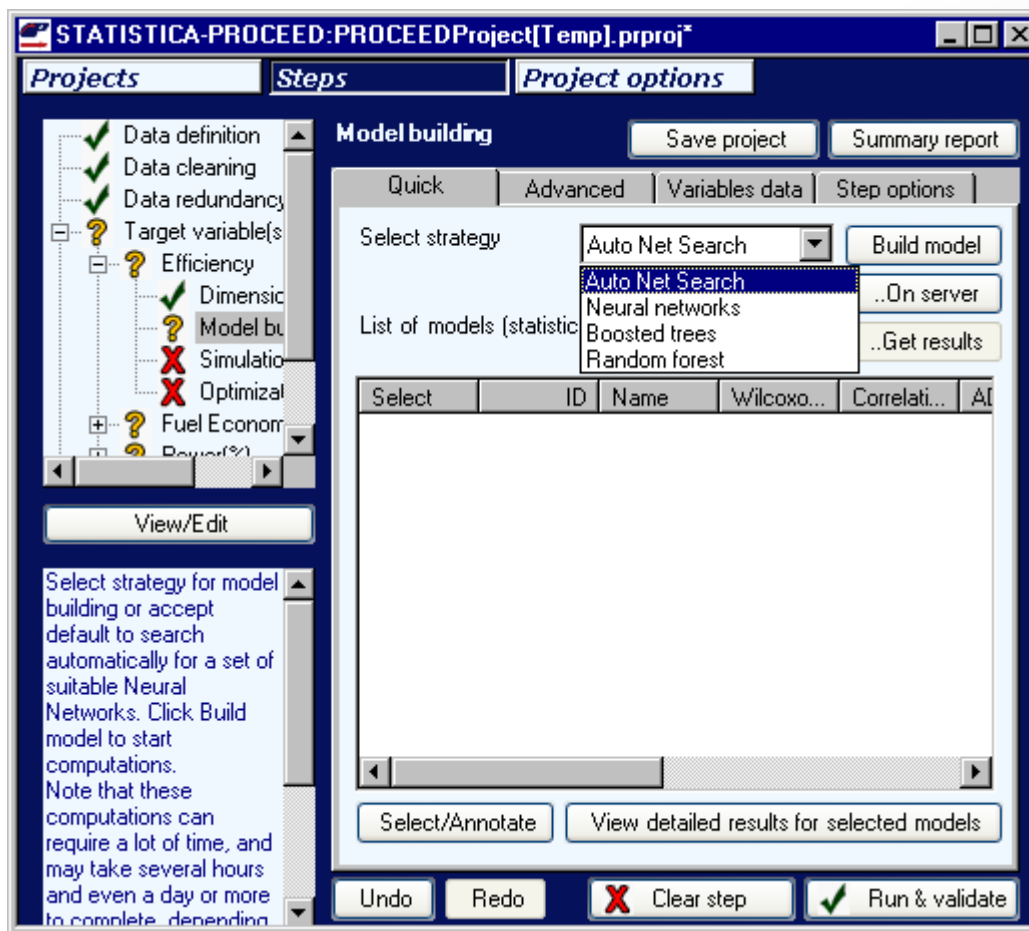


Największa odległość w sensie Mahalanobisa

Redukcja wymiarów



Budowa modelu



STATISTICA-PROCEED:PROCEEDProject[Temp].prproj*

Projects | **Steps** | **Project options**

Model building [Save project] [Summary report]

Quick | Advanced | Variables data | Step options

Select strategy: Auto Net Search [Build model]

List of models (statistic):
 Auto Net Search
 Neural networks
 Boosted trees
 Random forest

Select	ID	Name	Wilcoxo...	Correlati...	At

[Select/Annotate] [View detailed results for selected models]

[Undo] [Redo] [X] Clear step [✓] Run & validate

Select strategy for model building or accept default to search automatically for a set of suitable Neural Networks. Click Build model to start computations. Note that these computations can require a lot of time, and may take several hours and even a day or more to complete, depending

Symulacja

STATISTICA-PROCEED:PROCEEDProject[Temp].prproj*

Projects | **Steps** | **Project options**

Simulation Save project Summary report

Quick | Variables data | Step options

Variable Name	Distrib...	KS Sta...	KS P-v
Input01	Normal	0.0605	0.8251
Input02	LogNo...	0.0634	0.7834
Input13	Normal	0.0799	0.5173
Input33	Normal	0.0734	0.6232
Input36	Normal	0.0672	0.7253
Input45	LogNo...	0.0632	0.7870
Input49	Normal	0.0709	0.6653
Input51	LogNo...	0.0575	0.8745
Input59	Normal	0.0654	0.7536

Fit Edit Simulate

Model ID	KS Stati...	KS p-value	AD Statisti...	AD p-val
16				
20				
19				
18				

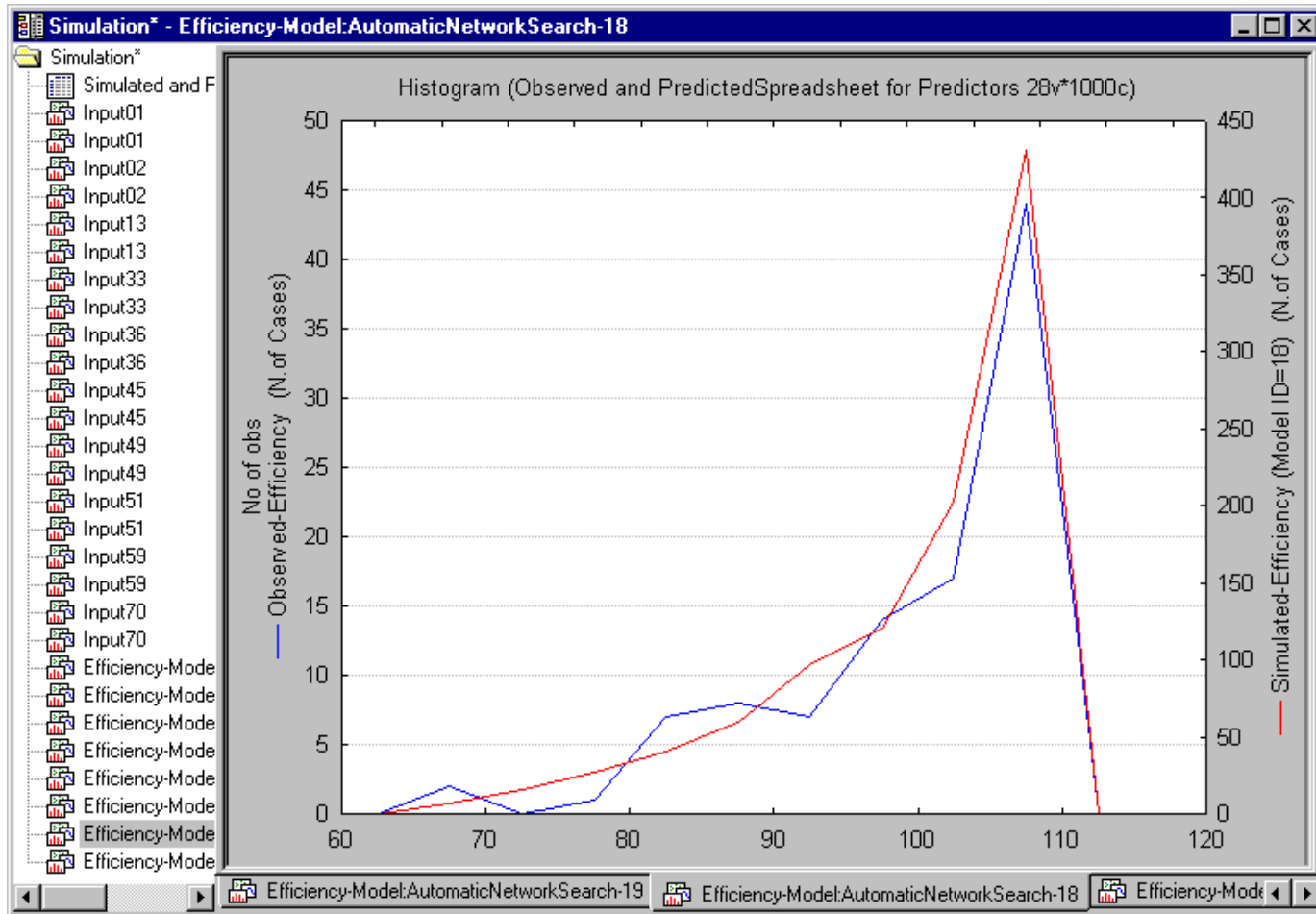
Model selected for final deployment (ID): **none**

Undo Redo X Clear step ✓ Run & validate

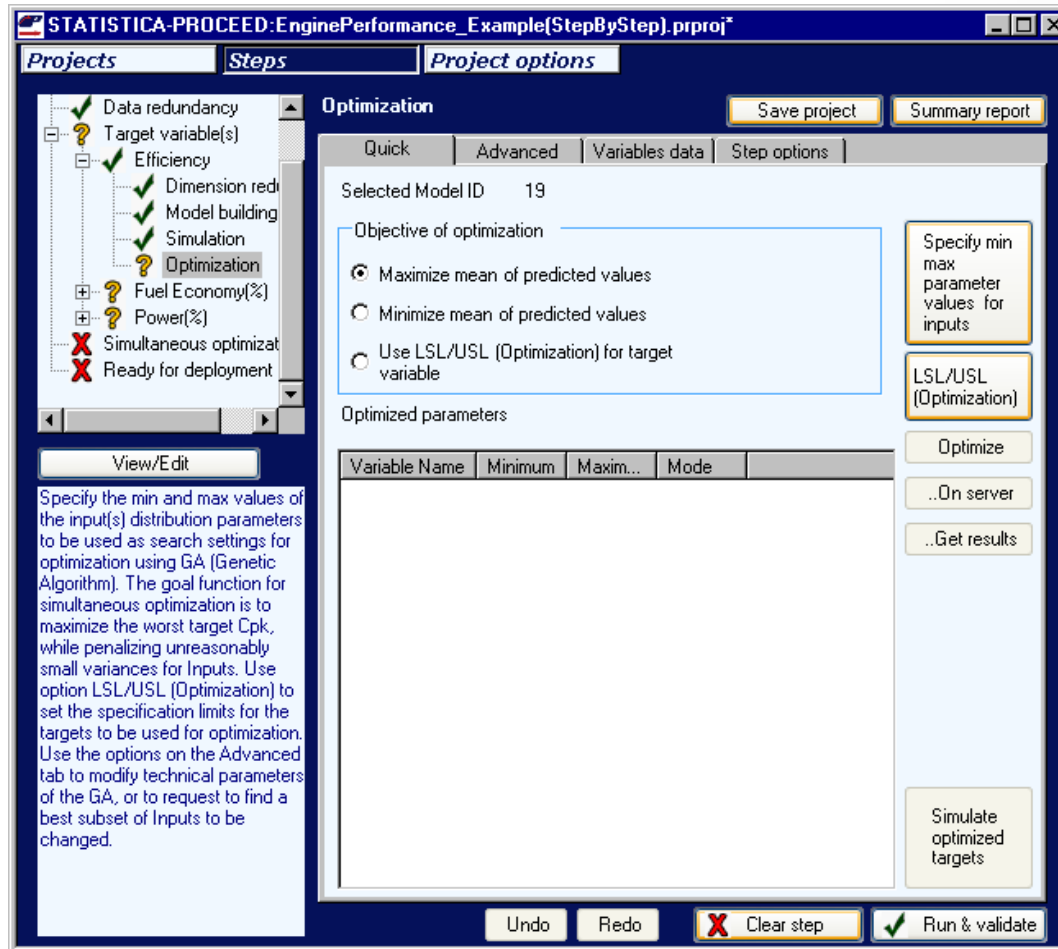
View/Edit

Simulate using selected distributions to create histograms of observed Target values and simulated values given the current models and distributions for the Inputs. Change the distributions for the Inputs if desired.

Wyniki symulacji



Optymalizacja



STATISTICA-PROCEED:EnginePerformance_Example(StepByStep).prproj

Steps

- ✓ Data redundancy
- ? Target variable(s)
- ✓ Efficiency
 - ✓ Dimension red...
 - ✓ Model building
 - ✓ Simulation
 - ? Optimization
- + ? Fuel Economy(%)
- + ? Power(%)
- ✗ Simultaneous optimizat...
- ✗ Ready for deployment

Optimization

Quick | Advanced | Variables data | Step options

Selected Model ID 19

Objective of optimization

- Maximize mean of predicted values
- Minimize mean of predicted values
- Use LSL/USL (Optimization) for target variable

Optimized parameters

Variable Name	Minimum	Maxim...	Mode

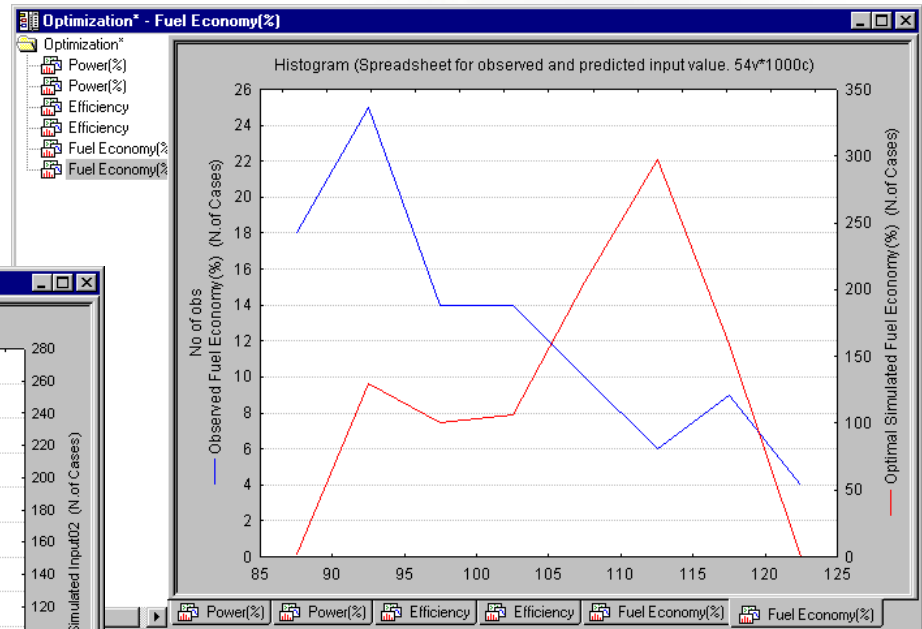
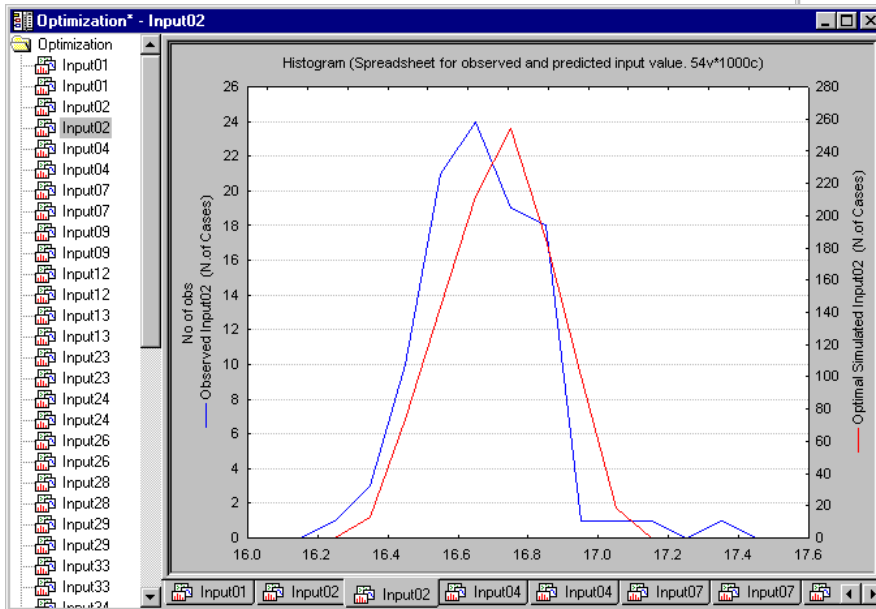
Specify the min and max values of the input(s) distribution parameters to be used as search settings for optimization using GA (Genetic Algorithm). The goal function for simultaneous optimization is to maximize the worst target Cpk, while penalizing unreasonably small variances for Inputs. Use option LSL/USL (Optimization) to set the specification limits for the targets to be used for optimization. Use the options on the Advanced tab to modify technical parameters of the GA, or to request to find a best subset of Inputs to be changed.

Buttons: Save project, Summary report, Specify min max parameter values for inputs, LSL/USL (Optimization), Optimize, ..On server, ..Get results, Simulate optimized targets, Undo, Redo, Clear step, Run & validate

Wyniki optymalizacji

Wyjście

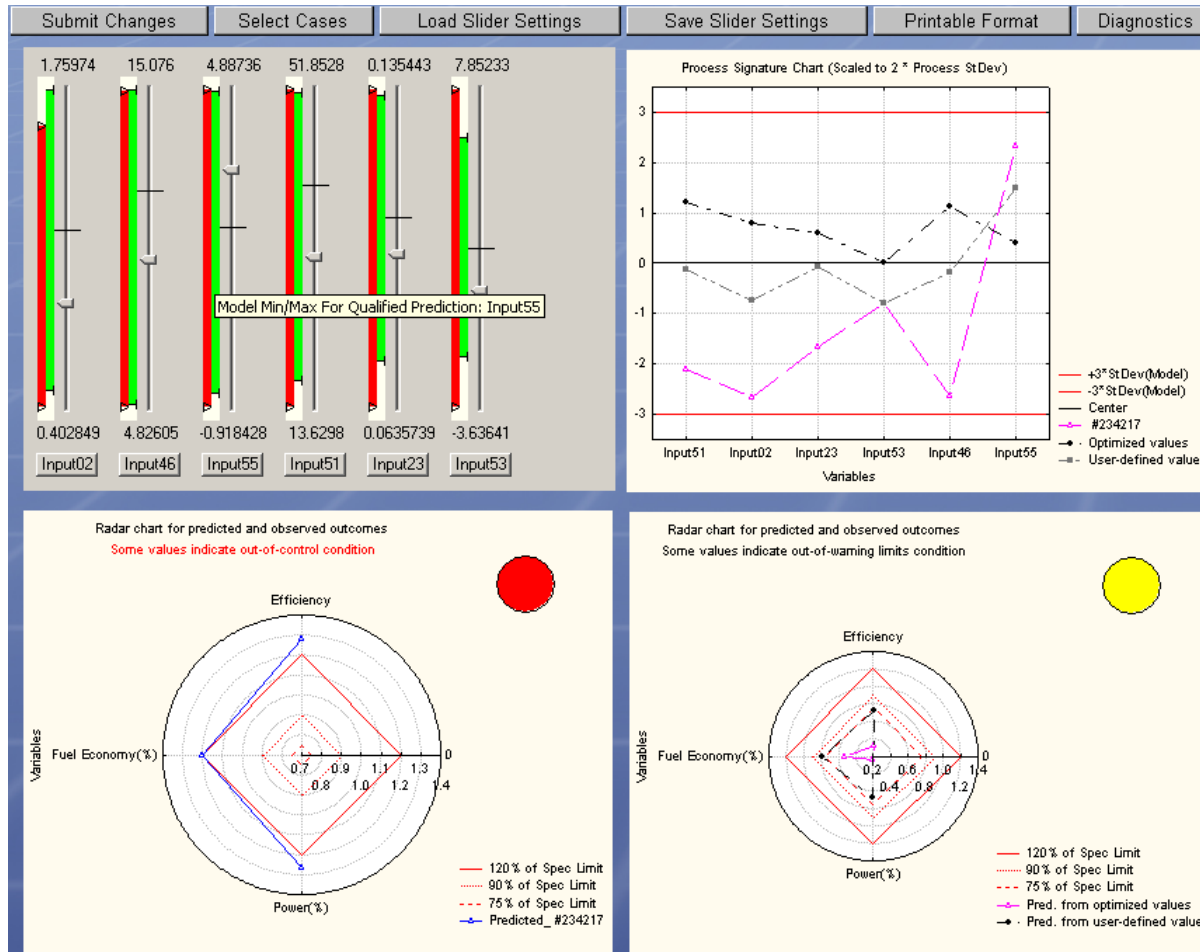
Wejście



Wdrażanie modelu

- ❑ Środowisko ADE (*Actionable Decision Environment*)
- ❑ Jak dane mówią nam co robić, czyli scenariusze *what-if*
- ❑ Przewidywanie konsekwencji zmian parametrów procesu produkcyjnego
- ❑ Przewidywanie jakości produktu przed testem

Wdrażanie modelu



Wdrażanie modelu

WebSTATISTICA: /Temporary/(dev)tom/This Session/Unnamed/SWMON.STA

File SEWSS Statistics Graphs Data My Menu Tools Help

Save Whatif Settings Printable Format Diagnostics

for Defined Values, ▾ - Point Estimate
 ned values.

Input23

Input53

Input46

Histogram of Simulated Target: Efficiency

Legend for Histogram:

- Fitted
- User-Def.
- Optimized
- Mean Observed
- Mean User-Def.
- Mean Optimized
- User-Def. Point
- Observed

Median; Box: 25%-75%; Whisker: Non-Outlier Range

Legend for Radar Chart:

- Median
- 25%-75%
- Non-Outlier Range
- Outliers
- Extremes

Podsumowanie

- ❑ Rozszerzenie Six Sigma - kolejny etap rozwoju jakości
- ❑ Skuteczne wydobywanie wiedzy i wdrażanie modeli w przedsiębiorstwie
- ❑ Return of Investment