대한산업공학회/한국경영과학회 2002 춘계공동학술대회
한국과학기술원(KAIST) 2002년 5월 3일-4일

Productivity Improvement by developing statistical Model

발표자 : Ill-Chul Shin
경상북도 포항시 남구 동촌동 1번지 POSCO 포항제철소 공정출하부
Jong-Hwan Park
경상북도 포항시 남구 동촌동 1번지 POSCO 포항제철소 스테인레스부

Abstract

POSCO #2 Stainless steel making plant produces more than 600 thousand ton per year with a variety of products consisting of austenite and ferrite stainless steel to meet customers' needs since 1996.

The plant has four different major processes, that are, EAF-AOD-VOD-CC to finally produce semi-product called as slab. In this study, we importantly took AOD process into consideration due to its roles such as to check and verify the final qualities through sampling inspection. But the lead-time from sampling to its verification takes five to ten minutes causing productivity loss as much as the lead-time as a result. Of all indices for quality and process control the plant has, carbon ingredient in liquid type of steel is the most important since it affects in a great way to the characteristics of steel, if any problem, customers not satisfied with quality could issue a claim; therefore there is no way but to guarantee it before delivery.

In this study, to reasonably reduce lead-time can save a cycle time and finally improve our productivity from a state-of-art alternative just such as applying statistical model based on multi-regression analysis into the A.O.D line by analyzing the statistical and technical relationship between carbon and the relevant some vital independent variables.

In consequence, the model with R-square 87% allowed the plant to predict, abbreviating the process in relations to sampling to verification, approximately the value of [C] so that operators could run the process line with reliability on data automatically calculated instead of actual inspection. In the future, we are going to do the best to share this type of methodology with other processes, if possible, to apply into them.

1. Introduction

1.1. Infrastructure

In many cases, especially in the area of steel

Terms		Specification	
		#1 SMP	#2 SMP
CAPACITY		90 TON/HEAT	90 TON/HEAT
VASSEL VOLUME		47 m3	56 m3
TOP LANCE	FLOW RATE	100 N m3/MIN	120 N m3/MIN
	GASES	O2	O2,N2,Ar
TUYERE	FLOW RATE	100 N m^3/MIN	100 N m^3/MIN
	GASES	O2,N2,Ar	O2,N2,Ar
REFRATORY		D/O+MgO-D/O	D/O+MgO-D/O

[Picture 1. Process Specifications]

industry we have been trying to get some useful information through extracting and analyzing data out of a variety of databases. The company POSCO has been building them up under state-of-art server-client computing environment to be able for users to easily gather data end-user want to do analysis and report useful information relevant to process control as well as business reports since the mids of 1990.

Without constructing computational infrastructure, it must be impossible, CEO of the company thought, to control and manage it, furthermore to pursue efficiency of management. Thanks to this activity, most of engineers can save their time and trials to get some data. Hereby I would like to emphasize the importance of computing environment, that is, on-line data gathering system based on DBMS.

In this study, we importantly took AOD process into consideration due to its roles such as to check and verify the final qualities through sampling inspection. But the lead-time from sampling to its verification takes five to ten minutes causing productivity loss as much as the lead-time as a result. Of all indices for quality and process control the plant has, Carbon ingredient in liquid type of steel is the most important since it affects in a great way to the characteristics of steel, if any problem, customers not satisfied with quality could issue a claim; therefore there is no way but to guarantee it before delivery.

1.2. Problems to solve

There were largely three different problems for instance, first, impossible was to predict the final carbon ingredient since no sensitive model was not developed, therefore operators in the line could use raw materials more than required or less which caused operating cost.

Second, Process delay due to verify the final chemical component such as carbon, manganese, nickel and the like, was inevitable which caused productivity loss.

More serious problem was that operator couldn't make an exact decision on the amount of oxygen blowing to control the final carbon ingredient. These are why we need to do this study.

2. Body

2.1. Data file Description

Like mentioned above, to gather some data from data base is very easy but all the data we required are not involved in it, for example the use frequency of furnace for which manual note is necessary. We collected more than 400 observation for the analysis during one month when the underlying process had been stable without any facility trouble. Before gathering data, we are to define critical process factors to the underlying quality characteristic summarized in the table1 [critical factors to the quality] by brain-and-storming discussion with process engineers. Selected are some input factors already decided before this AOD process and the process related factors like as the amount of oxygen blowing and sub material's amount and the like.

Items	Process Factors
Before process	Start carbon, steel amount, the use frequency of furnace, etc
In process	Slag volume, process temperature, the amount of oxygen blowing, etc

[Table1. Critical process factors to the quality characteristic]

2.2. Technical Outlier

Process engineers need to clarify if all the data is accurate and actual with technical viewpoint to prevent "garbage-in, garbage-out" phenomenon. Most of analysis procedure is sure based on sample data from population. Experiences with technical viewpoint tell useful data from garbage type of data.

2.3. Univariate Analysis

As the first step, we have done unviariate analysis to know its basic statistics like as mean, variance each variable and to help us understand each variable's distribution. Especially essential is to check normality of the underlying dependant variable since regression analysis is based on the assumption of normal distribution. For other independent variables, data usability that is, whether the variables conform to real situation comes out. All variables should be summarized like the table 3 to know some attribute each.

Items	value
Mean	1699.1
Med	1698
Mode	1698
Std.dev	20.07
vairnace	402.70
Kurtosis	0.31
Skewness	0.05
Range	114
Min	1634
Max	1748
Sum	246365
N	145

[table 3. An example of univariate analysis]

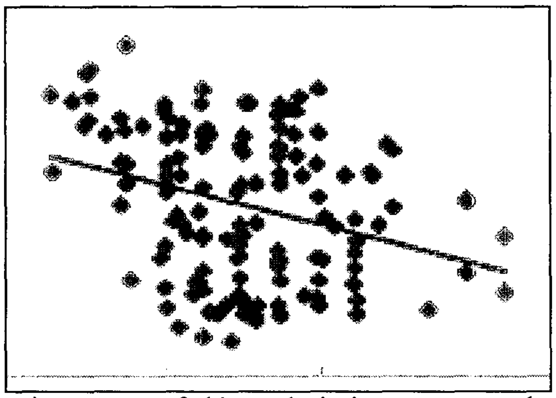
2.4. Correlation Analysis with Scatter-plot

In so many cases, process engineers and operators know already relationship between paired variables with their operating experience. But they didn't know exactly in a numeric way so as not to compare two different correlations. Also necessary is to draw scatter plot between the two to see if there is curve relation, say, nonlinearity.

Of course, all possible combination should be drawn and correlation coefficient calculated.

2.4.1. Between a Regressor and Predictors

A regressor, dependant variable can mainly be defined as quality characteristic and predictors as process factors that affect the regressor. The main purpose of this analysis is to compare the relationship between an independent variable and the dependant by selected independent variable. One can say, a predictor having higher correlation coefficient has strong relationship with regressor than a predictor having lower coefficient.



[picture 4. Oxygen blowing amount vs carbon]

The picture4 let us know the fact two variables, oxygen blowing amount and carbon, has negative correlation that is, carbon gets decreased as oxygen-blowing amount gets increased.

2.4.2. Between Predictors

We have to all the time consider multicorlinierity in the middle of executing regression technique to get some reasonable regression coefficient. That may be able to lead analysis result, if happen, to go seriously wrong way so that an analyst can not not admit technically and statistically. In case high correlation between the underlying two happens greater than its correlation coefficient 0.8, that means a strong possibility in that multicorlinierity exists later on. We made sure there is not any possibility of multi-collinearity from the correlation matrix but later time, it was required to check that again using VIF (Variance Inflation Factor) calculated.

2.5. T-test on H0: Parameter estimator = 0

All procedures of the regression analysis are required to run trial-and-error type of iteration because we are asked to re-define regression model over and over again following results from t-test when the null hypothesis is not rejected given a specific significant level, and sometimes from outliers. If p-value is lower than expected, we don't need to take a corresponding independent variable into consideration. Hereby a variable chcr_1_ was deleted as its p-value too low to reject the null hypothesis.

Variable	Label	DF	t Value	Pr > |t|
Intercept	Intercept	1	-2.9	0.0044
PAT	pat	1	-1.36	0.1756
AMT_G	amt_g	1	4.66	<.0001
sq1		1	14.8	<.0001
SLAG_AMT	slag_amt	1	2.62	0.0098
TEMP_1	temp_1	1	2.65	0.0092
CHCR_1	chcr_1	1	-1.29	0.1986
chcr_1_		1	0.25	0.8018
FESI_2	fesi_2	1	3.03	0.003

2.6. Regression Using Indicator Variables

Situation to use indicator variable would happen, especially in case the underlying model include a categorical variable. For example a variable has its value such as high-low, large-medium-small and so on.

2.6.1. Creating Indicator Variables

In case of this study, the furnace frequency can be technically divided into two different categories, new and old. Process engineers defined new furnace as its frequency less than or equal to 50 and old ladle as greater than 50 with technical viewpoints. Hereby in this regression model, we let "0" be new ladle and "1" the old.

2.6.2. Verifying Interaction effect

Some analysts emphasize the interaction between three variables or more would be ignored in reality because it may be impossible, if exists, to interpret and apply into the field. Interaction between the two can be considered in the regression model, which give us very useful information and guideline.

2.7. Statistical Outlier

Outlying some observation should be done very carefully as they, a data analyst and process engineers discuss enough to see if an observation is an outlier or not. Sometimes the wrong decision on an outlier can lead us to get a wrong conclusion statistically with a biased sample.

2.8. 1st Regression Model

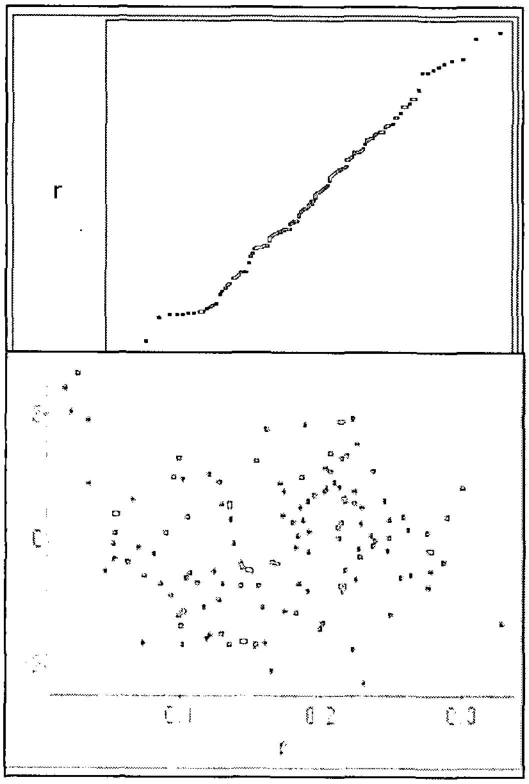
Through outlying 9 observations from the sample based on statistical criteria such as identifying influential cases or residual analysis, the first predicting model was built with 8 predictors (critical to process) of which R-square was 80%. But one of process engineers suggested that one important factor be ignored in the model technically; therefore, we revised the model once again that made the model more reliable.

2.8.1. Diagnosis on Error Term

Multiple regression model assumed that error term followed normal distribution with mean zero, variance σ^2. Therefore, we have to always diagnosis these pre-assumption using normality test and residual analysis.

2.9. Field Test for adapting

All of the statistical analyses and their results are fully influenced by a sample and furthermore no one can say this sample is completely split-image of the population. To test reliability of this model, the best way is to adopt it into the related process line. We had done it for one month before the model was retuned and actually adopted. The graph drawn on the left



side indicates how well predicted carbon value calculated by the regression model, fits actual carbon value.

2.10. Final Regression Model

The R-square became 87% after one more important factor added so that all process engineers and operators could trust its adaptability as well as usability. The description below implies what the final regression model looks like.

Parameter estimators in the table2. calculated by a statistical application software named SAS indicate regression coefficient.

3. Conclusion

The model with R-square 87% allowed the plant to predict, abbreviating the process in relations to sampling to verification, approximately the value of [C] so that operators could run the process line with reliability on data automatically calculated instead of actual inspection. Fitting rate as field criteria, defined as the difference between actual value and predicted value is as 96% high as operators can use. In the future, we are going to do the best to share this type of methodology with other processes, if possible, to apply into them. In order this model to be a sensitive to the process, it is very required to modify regression coefficients every three month so that process changes can be reflected and tuned. For doing so, we educated them, process engineers and operators, about how to do regression analysis, basic statistics and how to use the related software.

4. Further Study

In this study, to reasonably reduce lead-time can save a cycle time and finally improve our productivity from a state-of-art alternative. We believe also that other processes have similar situation to this process so as to adopt statistical techniques and do quality and productivity better done. We are also planning to develop a dynamic on-line system with database without manual operation such as data input for calculating the final carbon.