NEW ASSOCIATION IN BIO-S-POLYMER PROCESS

Long Flory
School of Business, Virginia Commonwealth University
Snead Hall, 301 W. Main Street, Richmond, VA 23284

ABSTRACT Small firms generally do not use designed experiments and statistical analyses that many large companies use to improve sustainability. This paper provides an example to small firms who wish to develop formidable capability over its competitors. The paper illustrates how a small firm leveraged designed experiments and statistical analyses to slash its production costs, enhance its organizational learning, and turn its problems to its advantage. This demonstration may encourage small firms to embrace statistical thinking and practice that derives competitive edge.

PART 1 INTRODUCTION

ABS Bio, Inc. is a small biotechnology company producing two staple products, Bio-S-Polymer and H2-E. Bio-S-Polymer is an effective waste water treatment medium. The production process of Bio-S-Polymer is summarized in **Figure 1**. During the production, the reaction vessels are sealed to control inside temperature, humidity and pressure in the specific ranges. Each of the reaction vessels produces one cubic foot of Bio-S-Polymer after a fixed amount of Bio-Bases reacts with the three fuels (fuel 1, fuel 2 and fuel 3) to formulate special molecule-chains called Bio-S-Chains of Bio-S-Polymer. The more Bio-S-Chains there are in one cubic foot of Bio-S-Polymer, the higher is the quality of Bio-S-Polymer. The number of Bio-S-Chains in one cubic foot of Bio-S-Polymer is defined by the quantities of the three fuels (quantity of fuel 1 = e1, quantity of fuel 2 = e2, quantity of fuel 3 = e3).

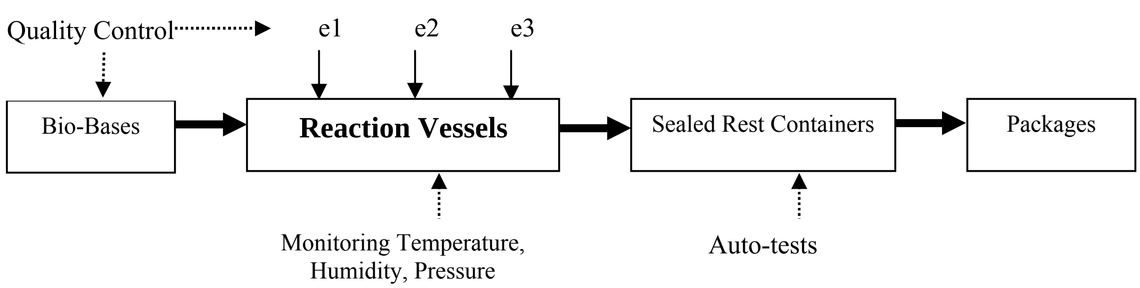


Figure 1: Bio-S-Polymer Production Process

In 2007, ABS Bio improved the reaction vessels of its Bio-S-Polymer process (Process-99). Unlike the old reaction vessels (99-Vessels), the new vessels (07-Vessels) reduce the variation of humidity and thus, increase the quality of Bio-S-Polymer. Unfortunately, 07-Vessels increase the production costs. To account for the increased production costs, the company implemented a 14% price increase to its Bio-S-Polymer. Yet, many existing customers were only willing to accept a price increase less than 8%, and ABS Bio lost a number of its customers. September 2008, ABS Bio teamed with Crystal Consulting, a business consulting firm, to seek ways to reduce the cost of the new Bio-S-Polymer process, Process-07.

After preliminary analysis, the managers in ABS Bio believed that reducing production costs of Process-07 required a new design and that the best solution was to roll back 99-Vessels while improving the design of Process-07. However, the consultant from Crystal Consulting proposed a possible alternative by questioning the perceived association between the input fuels and the output Bio-S-Chains in Process-07. In Process-07, ABS Bio used the same relationship between the output and the input as they did in Process-99. The relationship in Process-99 between the number of Bio-S-Chains in one cubic foot of Bio-S-Polymer and the quantities of the three fuels (e1, e2 and e3) is described by the empirical equation (1), Bio-S-Equation.

$$\hat{Y} = 39.8 + 7.6\,e1 + 1.07\,e2 + 0.65e3 \quad (1)$$

The consultant reasoned that Process-07 might change the relationship depicted by Bio-S-Equation. If the fuel-consumption is less in Process-07 than in Process-99, the production costs may be lowered considerably. ABS Bio accepted the proposal of the consultant, and commissioned a project team to conduct an experiment, Process-07 Experiment, to answer three key questions.

1. Does Bio-S-Equation best describe the relationship between the output Bio-S-Chains and the input e1, e2 and e3 in Process-07? (Hereinafter, "the output Bio-S-Chains" may present as "the output", and "the input e1, e2 and e3" as "the input".)

2. If the answer to question 1 is "No", what is the best empirical function that describes the relationship between the output and the input in Process-07?

3. Does the new relationship offer a reduction in the input, leading to a significant reduction in the production costs, so that the price of Bio-S-Polymer may meet the customers' expectation?

PART 2 DESIGNING PROCESS-07 EXPERIMENT

Part 2.1 Statement of Project Objectives

Establish an appropriate regression equation to describe the association between the output and the input in Process-07; examine whether the input can be reduced in achieving the quality standard of Process-07 and whether the reduction in the input can lower the price of Bio-S-Polymer as required by the customers. If all these can be done, new specifications and price policies will be established and implemented.

Part 2.2 Statement of Current Subject-Matter Knowledge

Bio-S-Equation was established via operational experience for Process-99. Since Process-07 reduces the variation of humidity inside the reaction vessels and improves the quality of B-S-Polymer, it is reasonable to perceive that Process-07 may establish a different relationship between the output and the input.

Part 2.3 Variables to Be Studied

Response Variable	**How Measured**
The number of Bio-S-Chains in one cubic foot of Bio-S-Polymer**:** The output Y	read auto-tests, use trillion as unit

Potential Predictor Variables	**Value of Selection**
1) The quantity of fuel 1 (e1)	see **Table 1,** use kg as unit
2) The quantity of fuel 2 (e2)	see **Table 1,** use kg as unit
3) The quantity of fuel 3 (e3)	see **Table 1,** use kg as unit

Background Variables	**How to Control or Measure**
1) Temperature	Controlled constantly at the specification
2) Pressure	Controlled constantly at the specification
3) Humidity	Controlled constantly at the specification
4) Quality of Bio-Bases	Controlled constantly at the specification
5) Quality of E1	Controlled constantly at the specification
6) Quality of E2	Controlled constantly at the specification
7) Quality of E3	Controlled constantly at the specification
8) Accuracy of testing equipment	Gauged to the same standard in each trial
9) Monitoring and testing methods	Automatic and consistent during the experiment
10) Operator	Two operators rotated, record operator's name
11) Process	One vessel controlled at standard conditions
12) Reaction duration	Production length: 145 minutes per trial

Because the production process is sealed from the outside environment, the time when trials were conducted does not have discernable effect on Y. So, time was not regarded as a background variable. The background variable, ɵOperator, may have discernable effect on the response variable Y. The team created a dummy variable EP (EP = 1 for one operator, EP = 0 for the other) for examining whether ɵOperator indeed had discernable effect on the sample output Y (see **Table 1**).

Part 2.4 Method of Observation and Randomization

The sample output Y (see **Table 1**) was observed in the twenty trials conducted in the twenty consecutive mornings, starting at 8:30am and lasting 145 minutes each morning. Randomly ordered by the computer, each of the twenty sets of e1, e2 and e3 (see **Table 1**) was used once. One of the two operators selected randomly by the computer operated the experimental Process-07 each morning.

Part 2.5 Generalization

The ranges of e1, e2 and e3 (see **Table 1**) include the possible values in the regular production. All the processing conditions in the experiment met Process-07 production standards.

Part 2.6 Design Matrix (see **Table 1**)

Part 2.7 Methods of Statistical Analyses

Any unusual observation of Y would be investigated and removed only when objective evidence indicated that the unusual observation did not represent the situation under study. The removal of an unusual observation should be recorded in the experiment documents. Graphical analyses of the sample data and residuals were conducted to examine visually the relationship between Y and e1, e2 and e3, assess the quality of the data and the validation of the assumptions, and probe possible ways to improve the-best-fit model. A thorough regression analysis was conducted to ensure that the final regression equation had a conceptual basis, fitted the sample data, and was free of any discernable deficiencies.

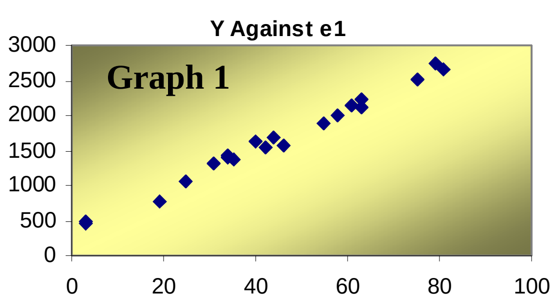
Table 1: Sample Data

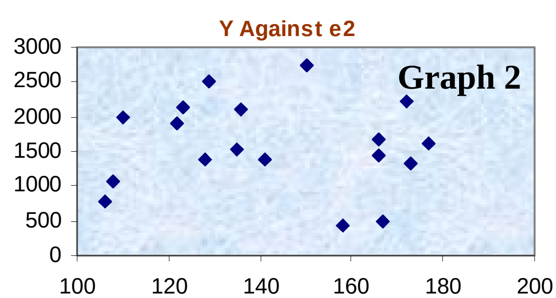
e1	e2	e3	Y	EP
34	166	149	1431	1
58	110	148	2004	0
3	158	156	443	0
63	136	163	2120	1
19	106	163	773	0
55	122	170	1890	1
35	128	176	1377	0
34	141	178	1388	0
44	166	183	1683	0
31	173	187	1327	1
42	135	189	1542	1
61	123	191	2135	0
81	97	196	2654	1
75	129	196	2521	0
63	172	208	2223	1
46	96	210	1571	0
79	150	216	2752	0
3	167	172	481	1
25	108	219	1064	0
40	177	223	1628	0

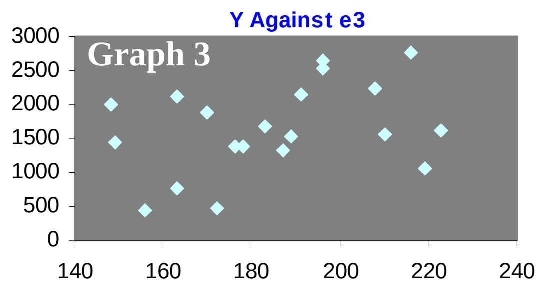
PART 3 GRAPHICAL AND INFERENTIAL ANALYSES

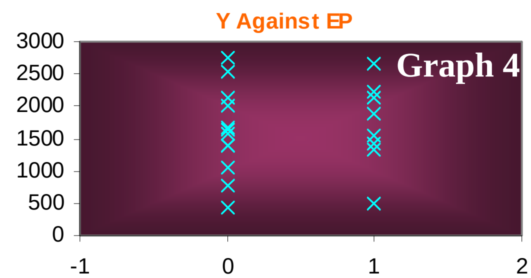
Part 3.1 Graphical Analyses

Graphs 1 ~ 4 provide visual reviews of the sample data Y against each of e1, e2, e3 and EP. The sample output Y is strongly related to e1 with a positive linear association. No appreciated relationship exists between Y and e2, as well as e3 or EP. No indication of outliers is in the graphs.









Part 3.2 Regression Analyses

The team started exploring the linear regression function with 11 terms – e1, e2, e3, EP, $(e1)^2$, $(e2)^2$, $(e3)^2$, e1*e2, e1*e3, e2*e3 and e1*e2*e3 – by using the Forward and Backward Stepwise procedure in Minitab. The procedure resulted in the least square equation, Function A (see **Table 2**), the evaluation to which follows.

Response is Y on 11 predictors, with N = 20 (Alpha-to-Enter: 0.05; Alpha-to-Remove: 0.05) Function A: $\hat{Y}$ = 21.56 + 29.64 e1 + 2.23 e2								
	T-values	**P-values of T-statistic**	Source	DF	SS	MS	F	P
e1	71.71	0.000	Regression	2	8173223	4086612	2663.70	0.000
e2	6.42	0.000	Residual Error	17	26081	1534		
R^2 = 99.68% R^2 (adj) = 99.64%			Total	19	8199305			
Source	DF	Seq SS						
e1	1	8109939	F(e1) = 8109939/1534 = 5286.79		P(e1) = 0.000			
e2	1	63284	F(e2 | e1) = 63284/1534 = 41.25		P(e2 | e1) = 0.000			

Table 2: Function A with e1 and e2

- Function A makes sense – in the ranges of e1 and e2 given in **Table 1**, if e1 or e2 has discernable effect on the output Y, the effect should be positive when everything else is held constant at the specific standard condition.

- The R^2 and adjusted-R^2 are almost perfect and identical (99.68% and 99.64%). The P-value of the F statistic is zero. So, Function A explains almost 100% of the variation in the sample output Y.

- Each of e1 and e2 has discernable incremental contribution to explaining the variation in the sample output Y after the effect of the other is accounted for. P-values of the T statistics are zero for both of the e1 and e2 coefficients, and P-values of the Partial F statistics, F (e1) and F (e2 | e1), are zero.

In order to provide convincing results, these questions need to be answered.

- Does e3 have any effect on the variation in the sample output Y?

- What is the individual effect of each e1 and e2 on the variation in the sample output Y?

- How much better is Function A than the other first-order linear functions involving e1, e2, e3 and EP?

To answer the questions, eight first-order linear equations were analyzed (see **Table 3**). The eight equations are designated as RE1, RE2, RE3, RE4, RE5, RE6, RE7 and RE8 respectively.

Regression Equation	MSE	F-value (F statistic)	P-value (F statistic)	T-value (T statistic)	P-value (T statistic)	Adj. R^2
$\hat{Y}$ =362+28.9e1 (RE1)	4965	1633.51	0.000	e1: 40.42	e1: 0.000	98.8%
$\hat{Y}$ =1722-0.59e2 (RE2)	439695	0.65	0.431	e2: -0.80	e2: 0.431	0.0%
$\hat{Y}$ =-224+10.2e3 (RE3)	398207	2.59	0.125	e3: 1.61	e3: 0.125	7.7%
$\hat{Y}$ =231+ 28.7e1+0.775e3 (RE4)	4942	821.06	0.000	e1: 37.86 e3: 1.04	e1: 0.000 e3: 0.313	98.9%
$\hat{Y}$ =406-4.60e2+10.2e3 (RE5)	404571	0.63	0.224	e2: -0.85 e3: 1.60	e2: 0.409 e3: 0.128	6.2%
$\hat{Y}$ =-56.6+29.5e1+2.19e2+0.501e3 (RE6)	1492	1826.43	0.000	e1: 67.78 e2: 6.35 e3: 1.22	e1: 0.000 e2: 0.000 e3: 0.243	99.7%
$\hat{Y}$ = 1612 + 97 EP (RE7)	453012	0.10	0.756	EP: 0.32	EP: 0.756	0.0%
$\hat{Y}$=5.6 +29.7e1+2.41e2-31.9 EP (RE8)	1351	2017.88	0.000	e1: 75.82 e2: 7.08 EP: -1.82	e1: 0.000 e2: 0.000 EP: 0.088	99.7%

Table 3: Eight First-order Linear Regression Equations

The statistics in **Table 3** provide the following insights on the relationship between the input and the output of Process-07 in the ranges of e1, e2 and e3 given in **Table 1.**

1. e1 by itself has discernable contribution to the variation in the sample output Y; for RE1, the P-value of the F statistic is zero, as well as the P-value of the T statistic of e1 coefficient. The adjusted R^2 of RE1 is very high 98.8%. e1 has discernable incremental effect on the sample output Y in the presence of e2 and/or e3; for RE4 and RE6, the P-values of the T statistics of e1 coefficients are zero.

2. In the presence of e2 and/or e3, e1 presents dominant influence on the sample output Y – For each of RE4 and RE6, the absolute value of the T statistic of e1 coefficient is significantly greater than that of either e2 or e3 coefficient.

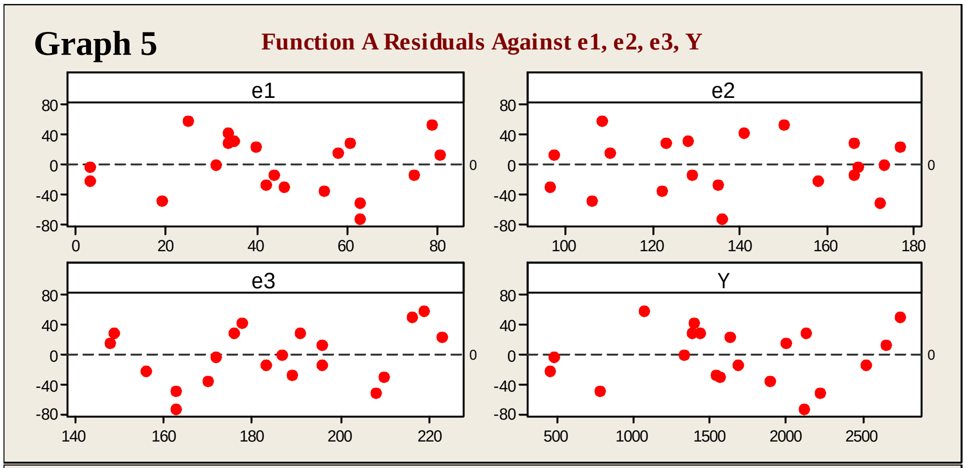
3. Neither e2 nor e3 has discernable individual contribution to the variation in the sample output Y; for RE2 and RE3, both P-values of the F statistics are greater than 0.05, as well as the P-value of the T statistic of e2 or e3 coefficient. e3 does not have discernable incremental effect on the sample output Y in the presence of e1 and/or e2 because in RE4, RE5 and RE6, the P-values of the T statistics of e3 coefficients are greater than 0.05. Similarly, e2 does not have discernable incremental influence on the sample output Y in the presence of e3, but does after the effect of e1 is taken into account.

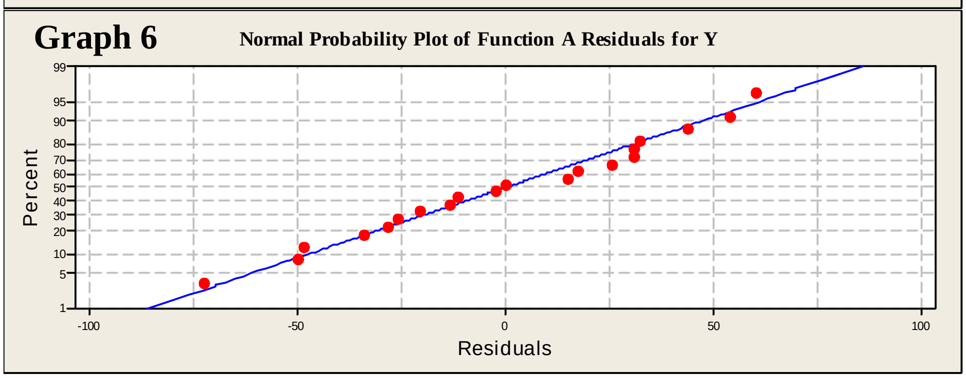
4. By the same token, EP has no discernable contribution to explaining the variation in the sample output Y either by itself or in the presence of e1 and e2, referring to RE7 and RE8.

Finally, Function A is obviously more efficient and effective to explain the relationship between the output and the input of Process-07 than any of the eight first-order functions in **Table 3**. Thus, the team selected Function A for the final evaluation.

Part 3.3 Examining Validation of Assumptions and Possible Improvement

Graph 5 displays the residuals of Function A against e1, e2, e3 and Y. **Graph 6** depicts the normal probability plot of the residuals of Function A.





The analysis of **Graphs 5 ~ 6** is preceded along with the inferences for Process-07 Experiment.

- There seems no additional relationship between the sample data Y and e1, as well as e2 or e3 except the association described by Function A because the residuals against e1, e2 and e3 spread randomly in horizontal bands centered round zero in general.

- The residuals against Y randomly reside in a horizontal band centered round zero, and the sample errors associated with the sample data Y distribute normally. So the assumption of constant error variance seems valid, and the assumption of normally distributed random errors is satisfied.

- The graphs indicate no appreciable outliers in the sample data Y.

- The assumption of independent random errors associated with the sample data Y seems valid; the residuals against e1, e2 and Y do not present appreciable patterns.

PART 4 CONCLUSIONS

Process-07 Experiment has discovered that in the ranges of e1, e2 and e3 given in **Table 1,**

- Bio-S-Equation is not appropriate to describe the relationship between the output and the input in Process-07.
- Function A named as Process-07 Equation – equation (2) hereinafter – is the best empirical equation to describe the relationship between the input and the output in Process-07.

$$\hat{Y} = 21.56 + 29.64\ e1 + 2.23\ e2 \qquad (2)$$

In Process-07 Equation, e1 has much more influence on the sample output Y than e2 does; the T-value of e1 coefficient is 71.71 versus that of e2 coefficient 6.42. The constant term 21.56 does not have practical meaning because Bio-Bases alone can not generate Bio-S-Chains. Process-07 Equation also reveals that in the ranges of e1, e2 and e3 given in **Table 1,**

- if e2 is held constant, the average number of Bio-S-Chains in one cubic foot of Bio-S-Polymer increases by about 29.64 trillions for every one kg increase of e1 (95% confidence).
- if e1 is held constant, the average number of Bio-S-Chains in one cubic foot of Bio-S-Polymer increases by about 2.23 trillions for every one kg increase of e2 (95% confidence).

In Process-07 production, Process-07 Equation can provide the estimate of not only the average but also the individual number of Bio-S-Chains in one cubic foot of Bio-S-Polymer for a specific set of input e1 and e2 from the ranges given in **Table 1.**

ABS Bio confirmed that the sample output Y from Process-07 Experiment met the product quality setup for Process-07 in terms of the combinations of e1, e2 and e3 given in **Table 1**.The comparison between Bio-S-Equation and Process-07 Equation provides these implications that are consistent with the fact of the quality improvement in Process-07.

- For the same sets of e1, e2 and e3 given in **Table 1**, the number of Bio-S-Chains in one cubic foot of Bio-S-Polymer produced by Process-07 is 1.33 ~ 2.99 times of that produced by Process-99 when everything else is held constant at the production standards.

- Every one kg increase of e1 in the range of e1 given in **Table 1** will result in about 22.04 trillions (22.04 = 29.64 – 7.6) more increase of the number of Bio-S-Chains in one cubic foot of Bio-S-Polymer in Process-07 than in Process-99 – when e2 and e3 are held constant at the levels given in **Table 1,** and everything else is held constant at the production standards.

- Every one kg increase of e2 in the range of e2 given in **Table 1** will result in about 1.16 trillions (1.16 = 2.23 – 1.07) more increase of the number of Bio-S-Chains in one cubic foot of Bio-S-Polymer in Process-07 than in Process-99 – when e1 and e3 are held constant at the levels given in **Table 1,** and everything else is held constant at the production standards.

Examining Process-07 Equation with the new data from the actual Process-07 production after Process-07 Experiment, ABS Bio has validated that Process-07 Equation best represents the relationship between the input and the output of Process-07, given the ranges of e1, e2 and e3 in **Table 1.**

PART 5 EXCEEDING PROJECT OBJECTIVES

The analyses given above indicate that in the ranges of e1, e2 and e3 given in **Table 1**, fuel 3 has no effect on the quality of Bio-S-Polymer in Process-07. Consequently, if the production runs within the ranges of e1, e2 and e3 given in **Table 1**, fuel 3 can be removed from the inputs of Process-07, and the total cost associated to e3 (about 16% of the total production costs) can be saved. The removal of fuel 3 leads to simpler production process and quality control, resulting in additional cost saving.

Using Process-07 Equation in production planning within the ranges of e1 and e2 given in **Table 1**, ABS Bio has reduced production costs of Process-07 about 20%, and therefore, has reduced the price of its Bio-S-Polymer up to 11%. ABS Bio exceeds the expectation of its customers. The company provides much higher quality of Bio-S-Polymer yet only 1% higher price than do its competitors. ABS Bio has used Process-07 Equation to establish new standards and marketing strategies for its Bio-S-Polymer since the regular production is within the ranges of e1 and e2 given in **Table 1**. Consequently, ABS Bio has not only won back the lost customers but also increased market share 3% and sales 17% despite the worldwide recession.

PART 6 FURTHER BENEFIT

1. Applying designed experiments and statistical analyses, ABS Bio has turned the risk from changing the reaction vessels in 2007 to great competitive advantage. The company has resolved the risk with only a fraction of the cost and the time required by a process redesign.

2. ABS Bio has developed a great appreciation for applying designed experiments and statistical analyses in problem solving.

3. The company has conducted a series of statistical experiments, using the other ranges of e1, e2 and e3, and has gained an in-depth knowledge about the relationship between the output and the input of Process-07. As a result, ABS Bio is more ready to meet the changes in market demand.

4. Through additional designed experiments and statistical analyses, ABS Bio has studied the effect of the temperature, humidity and pressure inside the reaction vessels of Process-07 on the quality of Bio-S-Polymer. The studies have revealed new ways to streamline the production process and continue reducing the production costs.

5. Learning from the Process-07 Experiment, ABS Bio has implemented a plan to excise good statistical quality control on Process-07. The plan includes training operators to improve their performance.

6. ABS Bio is on its way to develop an organizational culture of statistical thinking in problem solving and decision marking. The company has learned that statistical thinking improves the company's ability to anticipate market changes successfully and to gain advantage that is hard copied by its competitors.