STATISTICAL DESIGN OF EXPERIMENT

A TOOL FOR MINERAL ENGINEERS

H. PATNAIK

NML;JAMSHEDPUR

INTRODUCTION:

The classical methods of one-factor-at-a-time experiment(keeping all other remaining factors constant) requires a large number of trial involving time, energy, and money, yet the effect of interaction between various factors are not brought out clearly. In an effort to find optimum conditions with less number of experiments and to secure amount of quantitative information about the system, statistical or factorial design of experiments has been devised, A regression equation is developed:

$Yi=(x1,x2,...xi...xn)$

$Yi=\beta o+\Sigma\beta ixi+\Sigma\beta iixi^2+\Sigma\beta ijxixj$

$Yi=bo+\Sigma bixi+\Sigma biixi^2+\Sigma bijxixj$

where bo,bi,bij,bii are the best estimates(method of lowest squares) of regression coefficients βo,βi,βij,βii.

Efficiency of experimental investigation can be significantly improved by use of experimental design and data analysis based on statistical principles.

It is a mathematical method of drawing valid conclusion from a series of tests made in a predetermined pattern. It is a standard tool for the industrial experimenter as a chemical balance is for laboratory experimenter.

(+)

* Very helpful for preliminary investigation/screening experiment of systems with several independent variables. It effectively determines which factors are important so that these can then be more thoroughly investigated.

* Reduction in number of tests to be conducted(i.e. simultaneous rather than classical mwthod of varrying one by one)

* Experiments are well organised.

* Regression coefficients in conjunctions with existing knowledge or prior information can give better insight to the physical/ physicochemical phenomena occuring in the system and thus aid in quantifying the effects and interactions of various factors.

* Improvement in productivity and reliability of results.

Thus leading to the shortest path in optimization of the variables of the process, suitable for industrial research(more information at less cost and time)

FACTORIAL EXPERIMENTS:

By virtue of the literature already available or experience gained on similar systems, one decides about the various factors the range of variation(+1 for upper level, -1 for lower level, 'O'zero for base level) in coded form for each factor- the range chosen preferably being made narrow so that variation of the response variable is expected to be linear. While selecting the range of variation of the factor, first the base level is selected around which one wishes to vary the factor depending on prior information through literature or experience and then the interval or variation is selec

$x_j = z_j - z^{o}_j / \Delta z_j$

x_j=coded values, z_j, z^{o}_j are the actual or natural values for i th factor at any level, base level (0)

Δz_j=the range = $(z_{max} - z_{min})/2$, $z^{o}_j = (z_{max} + z_{min})/2$

x_j = coded value, takes values like +1, -1, 1.414, -1.414 etc.

Once the regression equation Yi=(Yi) is found, it is decoded to give the equation in which the effect of actual factors are related to response variables. (output)

There are various methods of finding regression rquations.

(I) YATES METHOD:

Besides giving the values of regression coefficients b,bij, it gives the additional information of SS (Sum of Squares), m.s(mean squares)

* Write the response in standard order (1,a,b,ab,c,ac,bc,abc) in a column. '1' denotes all factors are -ve: 'a' means factor A is a upper level (+1), all other factors being at lower level(-1)

and so on.

* Add these in pairs and then substract these in pairs(values of one ahead/succeeding from one behind proceeding in the table to form the next column.

* Continue similar operation one after another or number of columns equal to the number of factors k. Let the values in the last column last column be l.

* Actual effect = $l/2^{(k-1)}$

* b's = $l/2^{k}$

* Variance = SS = $l^2/2^{k}$, m.s = SS/df, SS = sum of squares, m.s = mean squares, df = degrees of freedom.

* Residuals = VT−ΣVi,VT(Total variance)=ΣY²−cF,cF(correction factor) correction factor = $(\Sigma Y)^2/2^{k}$

ΣY= the value in the last column against row marked '1'

ΣVi = Sumof variance due to all individuals and interactions

* F ratio = m.s /m.s residual, for a factor to be significant its F should be greater than F

n1=df corresponding to larger m.s

n2=df " " smaller estimate m.s (Denominator)

BULK FLOTATION OF COMPLEX Cu Pb Zn ORES

FLOTATION CONDITIONS (CONSTANT)

Addition	Quantity kg/ton	Conditioning time(min)
Sodium Silicate	1.5	5
Na2S	.5	2
CuSO4	[illegible]	5
Collector(varying)		3
Frother	.15	5

% Solids = 15, RPM = 1250

Factors:

	G	P	X
+	90	11	2.5
0	70	9	1.5
−	50	7	.5
Step	20	2	1.0

G=grind (%-200#)
P=pH of slurry fed,
X =collector (Now.IPX

X	Base	G	P	X	Reco.
1	+	+	+	+	92.1
2	+	+	−	+	91.25
3	+	−	+	+	74.11
4	+	−	−	+	64.63
5	+	+	+	−	88.67
6	+	+	−	−	88.38
7	+	−	+	−	61.14
8	+	−	−	−	55.69
9	+	0	0	0	82.98

YATES METHOD:

	1	2	3	$(Col3)^2/8$ =SS=ms,(df=1)	V.Ratio= ms/residual	Modified V.ratio
1	55.69	144.07	293.88	615.97		
a	88.38	149.81	322.09	104.83	1373.6661	943.45	94.1[illegible]
b	61.14	155.88	60.22	16.07	32.280612	22.17	[illegible]
ab	88.67	166.21	44.61	-13.79	23.770512	16.33	

615.97720.8* 723.08**

	1	2	3	$(Col3)^2/8$ =SS=ms,(df=1)	V.Ratio= ms/residual	Modified V.ratio
c	64.63	32.69	5.74	28.21	99.475512	68.32	6.81[illegible]
ac	91.25	27.53	10.33	-15.61	30.459012	20.92	
bc	74.11	26.62	-5.16	4.59	2.6335125	1.81	
abc	92.1	17.99	-8.63	-3.47	1.5051125	1.0	

	1	2	3	
add	255.57	359.26	354.68	664.84
even	360.4	361.54	368.4	71.96

615.97720.8 723.08**736.8

Note: a=G, b=P, c=X significant

* All interactions and residuals pooled together to 58.34 and this is used as modified residual for calculation of modified $F_{1,4,p=.05} = 4.54$

bj=(jy)/N=(jy)/8

	G	P	X
bj	13.13	2	3.52
T	20.00	2	1.00
steps			
bjT	262.6	4	3.52
K	constant		35.2
T'	7.5	.1	.1
real new step	7.5	.1	.1 (fixed)

limiting factor
0.1 kg/t, collector=x

Matrix II

Test	G	P	X	Recovery
10 (base)	70 7.5	9 .1	1.5 .1	82.98
11	77.5	9.1	1.6	85.49
12	85.0	9.2	1.7	92.68
13	92.5	9.3	1.8	92.23

Test 12 or 13 taken as optimum.

X test:

R=bo+b1G+b2P+b3X
=76.99+13.13G+2P+3.52X

Test run	G	P	X	Robs	Rcalc	X=(Robs-Rcalc)/Rcalc
	+	+	+	92.1	95.64	.1310
	+	-	+	91.25	91.64	.0016
	-	+	+	74.11	69.38	.3224
	-	-	+	64.63	65.38	.0086
	+	+	-	88.67	88.6	.0000
	+	-	-	88.38	84.6	.0005
	-	+	-	61.14	62.34	.0231
	-	-	-	55.69	58.38	.1239
						.6111

<X 7, 95%=14.07

Thus it can be said with confidence that the result of matrix I are represented by above linear equation.

IInd method:

factors / expt.No.	x0	x1	x2	...	x1x2	x1x3	...	x1x2x3	yi
1									
2									
3									
4									
5									
6									

$\Sigma x_{ij} y_i = (jy) = (\Sigma - 2(\Sigma - ve) = ()$

$\Sigma = (oy)$ = sum with all values +ve

$\Sigma(-ve)$ = sum of all -ve values only

(jy) = the +ve & -ve values in j th column (variable or factor) multiplied with corresponding yi values & then summed together.

$\Sigma x_{ij}^2 = (j)^2 = N$

$b_j = \Sigma x_{ij} y_i / \Sigma x_{ij}^2 = (jy)/(j)^2 = (jy)/N$

$S^2 b_j = Se^2 / \Sigma x_{ij}^2 = Se^2/(j)^2 = Se^2/N$

$Se^2 = \Sigma (y_i - y)/n-1$ = error mean square = replicate observations (n) at base level.

$t_j = [b_j / Sb_j] > t$ table at $p = .05$, $df = n-1$ ... to be a significant coefficient in regression equation.

Subscript j	0	1	2	...	12	13	...	123
(jy) or ()								
bj								
tj								

collect all significant coefficients to form a regression equation of $\hat{y}$ (estimate)

$\hat{y} = b_0 + b_1 x_1 + \ldots + b_{12} x_1 x_2 + \ldots$

$S_r^2 = \sum_{j=1}^{N} (y_j{}^2 - \hat{y})^2 / (N-l)$ = residual variance, l= number of significant coefficients

y_j = observed, $\hat{y}$ = calculated from regression equation.

F-Test: (test for adequacy of fit): The test is done for confirming reproducibility of results. The purpose is to show that there exists no significant variations from batch to batch at 5% confidence intervals due to personel or experimental errors.

$F = S_r^2 / S_e^2 < F \mid f_1 = N-l,\ f_2 = n-2$, n= number of observations at base level. So $\hat{y}$ adequately fits and is then converted [illegible]) to natural scale.

Example:

Factor	Code	levels		
		upper	base	lower
		+	Ø	−
%Bentonite	A	8.707	8.0	7.293
%Water	B	3.354	3.0	2.646

Test No.	Xo	A	B	AB	Yg
1	+	+	+	+	10.11
2	+	−	+	−	6.31
3	+	+	−	−	11.27
4	+	−	−	+	7.24
(jy)	34.93	7.83	−2.09	.23	

N=4

(AY)= [Σ−2(−ve)]
    = 34.93−2(6.31+7.24)= 7.83
(BY)= 34.93−2(11.27+7.24)= −2.09
(ABY)= 34.93−2(6.31+11.27)= −.23

j	0	A	B	AB
(jy)	34.93	7.83	−2.09	−.23
bj=(jy)/4	8.73 ✓	1.96 ✓	−0.52 ✓	.06
|Δbj|		where bj > |Δbj|		

Experiment at base level(Ø):

Test No.	A	B	Yg
5	Ø	Ø	9.64
6	Ø	Ø	9.40
7	Ø	Ø	9.52
8	Ø	Ø	9.24
9	Ø	Ø	9.33

$Se^2 = \frac{\sigma^2}{n-1}\Big|_0 = .02478$

$Sbj^2 = Se^2/(j)^2 = Se^2/N = .02478/4$

Sbj= 0.07871

In order that bjj is to be significant, tj=bj/Sbj>t table|p=.05

df=n-1=5-1=4

t table=2.78 at p=.05, df= n-1=5-1=4

bj>t table x Sbj

$|\Delta bj|$ = t table x S bj = 2.75 × .07871 ≈ .2188

$bj > |\Delta bj|$

CENTRAL COMPOSITE ROTATABLE DESIGN (C C R D)

A second order Orthogonal design is not rotatable and as a consequence the errors in y at experimental points on the response surface may be smaller than they are as determined from regression equation.

To make centre composite rotatable , $\alpha = 2^{k/4}$ or $2^{(k-1)/4}$ and n0 observations at base levels increased.

As before, in addition to 2 experiments, some more experiments are performed at ±α points at base levels.

STEPS:

* Find (0y), (1y), (2y) ....,(12y),....,(11y),... where '0' corresponds to x_0 where all x are +ve i.e +1
* (Σjjy)= (11y)+ (22y)
* Find a_1, a_2, a_3, a_4,...., a7 from Table- 1
* Find b_j, b_{ij}, b_{jj}, S_{bj}^2, S_{bij}^2, S_{bjj}^2, ±Δb according to equation

b	S_b^2	S_b	$\pm \Delta b = t S_b$
$b_0 = a1(0y) - a_2 \Sigma(jjy)$	$S_{b0}^2 = a1 S_e^2$	S_{b0}	$\pm \Delta b_0$
$b_j = a3(0y)$	$S_{bj}^2 = a3 S_e^2$	S_{bj}	$\pm \Delta b_j$
$b_{ij} = a4\ (ijy)$	$S_{bij}^2 = a4 S_e^2$	S_{bij}	$\pm \Delta b_{ij}$
$b_{jj} = a5(jjy) + a6\ (\Sigma jjy) - a2\ (0y)$	$S_{bjj}^2 = a7 S_e^2$	S_{bjj}	$\pm \Delta b_{jj}$

+a5(jjy)+L , where L = a6Σ(jjy) − a2(0y), a constant.

$b0 = AN^{-1}[2\lambda^2(k+2)(0y) - 2\lambda C\ \Sigma(jjy)] = a1(0y) - a2\ \Sigma(jjy)$

$b_j = CN^{-1}(jy) = a3(jy)$

$b_{ij} = \lambda^{-1} N^{-1} C^2(ijy) = a4\ (ijy)$

$b_{jj} = AN^{-1}\ [((k+2)\lambda - k)\ C^2\ (jjy) + (1-\lambda)\ C^2\Sigma(jjy) - 2\lambda C\ (0y)]$

$= a_5(jjy) + a_6\Sigma(jjy) - a_2(0y)$

Where $A = [2\lambda((k+2)\lambda - k)]^{-1}$, $C = N/\ \Sigma x_{iu}^2$

$\lambda = k(n_1 + n_2)/[(k+2)n_2] = kN/[(k+2)n_2]$

k = number of independent variables,

n_1 = number of centre points

n_2 = number of peripheral points

$N = n_1 + n_2$

No of Variable k	No of points in the 2k factorial	Star(+/-L) points 2k	Centre (base level)	Total N	L value 2k/4
2	4	4	5	13	1.414
3	8	6	6	20	1.682
4	16	8	7	31	2.0
5	32	10	8	50	2.378

No.of variable k	a1	a2	a3	a4	a5	a6	a7
2	0.2	0.1	0.125	0.25	0.125	0.01875	0.1438
3	0.16638	0.056791	0.073224	0.125	0.0625	0.006889	0.0695
4	0.142857	0.035714	0.041667	0.0625	0.03125	0.00372	0.035
5	0.0988	0.0191	0.0231	0.0312	0.0156	0.0014	0.0172

Example:

Using the same example as was considered earlier for 2k experiment but in this ±α=1.414 was added, resulting in four more experiments at base levels were conducted.

	LEVELS				
	+1	0	−1	±α	
				1.414	−1.414
Bent. x1	8.707	8.0	7.243	9	7
Moist.x2	3.354	3.0	2.645	3.5	2.5

+1, −1, 0, +1.414, −1.414 denote levelst(upper, lower,base, ±α),α=1.414] at which experiments were conducted.

No.	xo	x1	x2	x12	yg		
1	+	+	+	+	10.11		
2	+	−	+	−	6.31		
3	+	+	−	−	11.27	34.93	
4	+	−	−	+	7.24		
5	+	1.414	0		10.04		
6	+	−1.414	0		6.52		
7	+	0	1.414		7.05		
8	+	0	−1.414		8.82		
9	+	0	0		9.64		
10	+	0	0		9.4		114.53
11	+	0	0		9.52	79.6	
12	+	0	0		9.24		
13	+	0	0		9.33		

(oy) (1y) (2y) (12y)

oy	1y	2y	11y	22y	12y
114.53	12.82	-4.55	68.07	66.73	-.23

Σ(iiy)=134.8

bo=a1(oy) - a2Σ(iiy)= .2(114.53)- .1(134.8)= 9.426=bo

bi= a3(iy)= .125(iy)= .125(12.82) = 1.6025 = b1

.125(-4.55)= -.56875= b2

bij= b12=a4(ijy)=.25(-.23)=-.0575= b12

bii= a5(iiy)+a6Σ(iiy)-a2(oy)

.1251(iiy)+.0187(134.8)-.1(114.53)

.1251(68.07)-8.93224 = .416683=b11

.1251(66.73)-8.93224=-.584317=b22

t p=.05, y=2.776, δ² n-1|o = s²g or s²e= .0247799

s² bo=a1 s² y=.2

s² bi= a3 s² y=.125

s² bij=a4 s² y=.25

s² bijj= a7s² y=.1438

s² y	√s² bj=sbj	±tob=tsb
.0247799	.703987	.19543
	.055655	.154498
	.0787081	.2185
	.0596937	.1657

bo	b1	b2	b11	b22	b12	± bo	± bi	± bij
9.43	1.6	-.57	-.4167	-.584	-.057	.1954	.154	.218

neglecting b's <± bi, ie. only b12

yg= 9.43 + 1.60x1- .57x2- .42x²1- .59x²2

Summary:

* Statistical Design of Experiment is an efficient method with least number of experiments in a planned manner, one can find out the predominant effect of certain variables (xi) & for their interactions (xixj) which is not possible in classical experiment.

* First build up Full/fractional factorial experiment to screen out important variables & interaction by t-test (significant coefficients) & F-test (adequacy of fit)

* Use method of steepest ascent when first order coefficients are significant.

* Otherwise continue some more tests at star points $\pm\alpha$ and use CCRD for 2nd order regression.

BIBLIOGRAPHY

1. Davies OL - Design and Analysis of Industrial Experiments, ICI Ltd, Longman group Ltd; NY, London. 2nd ed(1978)

2. Box & Wilson KP, J of Royal Society.Vol 13, 1, (1951)

3. Chew Victor, Experimental Design in Industry, John Wiley NY (1958)

4. Cochran WG & Cox GM,'Experimental Design, Wiley NY (1950)

5. Kafarov, Cybernetic methods in analysis of Chemical Engineering Process, Mir Publishers (1976)

6. Fuersteraw MC, Flotation vol 2, AM Gaudin Memorial Volume (AL Popular p429), NY. AIME 1976.

7. Biles William & Swain JJ, Optimization and Industrial Experiment Wiley Inter Science (1980)