Effects of Culture Conditions on Growth of Kefir Grains using Milk for Nutrition and Health: An Application of Response Surface Methodoly with Box-behnken Design.

Edwine B. Atitwa [a]*, Joseph.K.Koske author [b], John M.Muindi [c]

[a]Department of Computer and Statistics, Moi University, P.O. Box 3900, Eldoret, and 30100, Kenya, Address,
[b] Department of Computer and Statistics,Moi University, P.O. Box 3900, Eldoret, and 30100, Kenya
[c]Department of Computer and Statistics,Moi University, P.O. Box 3900, Eldoret, and 30100, Kenya
[a]*Email: eatitwa@gmail.com*
[b]*Email: Koske4@yahoo.co.uk*
[c] *Email:johnkasome@yahoo.com*

Abstract

This combined process was successfully modeled by response surface methodology method with a Box–Behnken design. In this work, Box-Behnken Design of response surface methodology was used to investigate the effects of fat content in milk, incubation time and the number of rotations or shaking as culture conditions to for growing kefir grains using milk as culture liquid. The implementation of first order factorial experiment was based on a 3^k Box- Benhnken experimental factorial design with one replicate. Response surface methodology was adopted to express the output parameter (Size of the kefir grain) that is decided by the input parameters (Number of rotations, Time and milk fat content) which yielded the response surface models. The effects of the culture conditions on growth of kefir grains using milk was studied by use of first order model. The adequacy of the first order model was examined by use residual analysis, the normal plot, the main affects plots, the contour plot, and ANOVA statistics F-test, t-test, R^2, and the adjusted R. The optimization of culture conditions was achieved by the help of the steepest ascent method to locate the optimal domain and direction of growth of kefir grains. The first order model was predicted using a D-optimal criterion which was used to evaluate the growth of kefir grains. The confidence level to check the growth of kefir grain was at 95%. The data was analyzed using R statistical software and excel. Data was presented using tables and graphs. Milk fat content and number of rotations per minute had a positive effects while time had a negative effects on the growth of kefir grains using milk.

Keywords: Kefir grains; Response Surface Methodology; Box-behnken design; Culture conditions

Introduction (use bold for main headings like this one. do not use italic)

This paper focuses on the analysis of effects of milk fat content,number of rotation,and time on growth of kefir grains using milk based on response surface methodology with box-behnken design.

Background of the study

Kefir grains are combinations of yeasts and bacteria living on a substrate made up of a variety of dairy components (A Lewis, 2012). Furthermore, probiotic bacteria in kefir grains have shown many health-promoting properties, such as inhibition of detrimental effect of pathogens, enhancement of the intestinal barrier and modulation of the immune response. These health benefits have been observed both *in vitro* and *in vivo* experiments (Courtesy etal., 2007). **It can help prevent side effects and damage done by antibiotics; it may help with lactose intolerance (Dominic etal., 2015).** This study used RSM with BBD to develop the design for data processing. When treatments are from a continuous range of values, then a RSM which is a collection of mathematical and statistical techniques is useful for the modeling and analysis of problems in which a response of interest is influenced by several variables and the objective is to optimize this response (Montgomery 2005). first-order model with k independent variables can be expressed as. For this study we used BBD to study quantitative variables since all the factors are quantitative in nature and ratatability of the design was assessed.

1.2 Problem statement

The number of people feeding on chemically processed foods has drastically increased in Kenya and it leads to continue eating junk foods in excess, which can lead to obesity, cancer and other health problems. These foods have additive such as flavors, flavor enhancers, binders, color, fillers, preservatives, stabilizers, emulsifiers, etc. Some of these addictives have a high glycenic index which is a major contributor to diabetic (Ashley Pruett, 2010). Presently, many studies has used the direct search methods that find a local maximum moving on a function over the relative local gradient directions and the indirect methods that usually find the local ends by solving a set of non-linear equations, resultant of equaling the gradient from the object function to zero, i.e., by means of multidimensional generalization of the notion of the function's extreme points from elementary calculus give a smooth function without restrictions to find a possible maximum which is to be restricted to those points whose slope is zero in all directions. Considering the diversity in composition of the natural biomass in kefir grain production, as well as the structure and physicochemical properties of these probiotics, a universal processing protocol is not conceivable, and specific processes must be designed for each growth size depending on different factors and the culture composition used (Molina S etal., 2013). For such programs to be effective, it is essential that the models are accurate. Thus the main problem was to make the growth of kefir grains that yields grains that contain high level of probiotics using RSM with BBD techniques and checking the effects of milk fat content, number of rotations and time during production that yields kefir grains with high level of probiotics that are healthy and nutritious.

1.3 Purpose of the study

One of the component of Kefir grains is lactic acid bacteria which have important applications (Welman and Maddox, 2003), including as thickeners, stabilizers, emulsifiers, pharmaceutical and chemical products (Palahikumar k etal., 2008).This component is important to stabilizes the effects of GMO and additive foods in human body to provide good nutrition and better health for all people and especially those suffering from diabetic and other related diseases. This study was done by carrying out experimentation to find out how milk fat content (grams), number of rotations (rpm), and Incubation time (hr) affects growth of kefir grains using milk to grow kefir grains with high yield of probiotics which is good for nutrition and health.

2.0 Litrature Review

The relationship between the response variable *y* and independent variables is usually unknown. And this requires the low-order polynomial model is used to describe the response surface *f*. This is used to meet results of understanding the behavior of response y on the independent variables $x_1, x_2, x_3, ...x_k$ (Pengpeng qiu, 2014).The method of steepest ascent given by Box and Wilson (1951) is a procedure for moving sequentially in the direction of the maximum increase in the response of variables. To overcome the scale dependence of the path of steepest ascent(Rodriguez etal.,2010) developed an adapted path of steepest ascent (ASA) which was used in this study to assess the suitability of the first order model.

This design was developed by Box and Behnken (Box Gep etal.2005) it provides three levels for each factor and consists of a particular subset of the factorial combinations from the 3^k factorial design. The actual construction of such a design is described in the three RSM books (Box and Draper, 2007).The use of the Box–Behnken design is popular in industrial research because it is an economical design and requires only three levels for each factor where the settings are −1, 0, and 1, since the values of the independent variables are all quantitative(M. Cavazzuti, 2013).

D-optimality (Kiefer J, 1974) was the first alphabetical optimality criterion to be developed. And it was used to study the optimal criterion of the first order model for studying the effects of culture conditions on growth of kefir grains using milk. The best design is the one with the highest D-efficiency. The D-efficiency of the standard fractional factorial is 100 % , but since it's not possible to achieve 100 % D-efficiency when pure quadratic terms are included in the model, it leads to some being less than 100%. And this is managed by running the order of the design runs randomly (W.A. Diamond etal., 2003) to find the ranks that achieve 100%.

3.0 Materials and methodology

3.1 Culture liquid and conditions

The culture liquid that was adopted to establish the effects was dairy milk .One spoon of the culture media(kefir grain) was stirred in to the milk and shaken for a number of times. Then the solution was kept at room temperature and controlled time. After that the grains were strained from milk which gave milk kefir that is rich in probiotics and nutritious for consumption that promote the biomass growth rate which was measured using the formula below

$$\%G = \left(\frac{X_{n+\frac{1}{2}} - X_n}{X_n} \right) \times 100 \tag{1}$$

Where:

G=Growth rate; X_n =Biomass weight after n days (grams); $X_{n+\frac{1}{2}}$ =biomass weight after $\left(n + \frac{1}{2}\right)$ days (grams)

All the results were carried out 3 times and the average was taken (Joe leech, 2015).

3.2 Box-bohnken design

The design was formed by combining three-level factorial design using the incomplete block of BIB with one associated class as shown below;

$$X = \begin{bmatrix} \pm 1 & \pm 1 & 0 \\ \pm 1 & 0 & \pm 1 \\ 0 & \pm 1 & \pm 1 \\ 0 & 0 & 0 \end{bmatrix} \quad (2)$$

Source (G.E.P Box and D.W. Behnken, 1960).

The above distinct runs satisfied the minimum distinct design points for any experiment of the given k parameters which is given by at least number of distinct points required for the experiment; N= $1+2k+\frac{k(k-1)}{2}$ (3)

Where K is the number of independent variables used in the experiment.

For data collection and processing, the factors in this study were assigned different values for the levels of each factor in reference with the minimum conditions of the other studies that have been carried out on growing kefir grain as shown in the table below for the three levels i.e. low(-1),medium(0) and high(+1).

Table 3.1: Table of conditions and their level of measurements for the Box-Bohnken experimental design.

Conditions	Measurement of levels		
	Low(-)	Medium(0)	High(+)
Time(hours)	24	36	48
Fat content(grams)	3.5	7.0	10.5
Shakers/minute	50	100	150

To make the factors to be dimensionless for the BBD, codification of the levels of the variables was done by using $x_i = \frac{X_i - X_0}{\Delta X_i}$ (4)

Where x_i is the dimensionless value of an independent variable; X_i is the real value of an independent variable; X_0 is the real value of an independent variable at the center point and ΔX_i is the step change (Tanyildizi MS *e tal.*, 2005).

3.3 First-order model

The first-order model was used to analyze the response surface and it gave a linear function of independent variables that explained the effects of the independent variable. The model generated was expressed as follows;

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon \quad (5)$$

The model was further manipulated by including interaction terms with three predictors and interaction terms as follows

$$= \beta_0 + \sum_{i=1}^{3} \beta_j x_{ij} + \varepsilon_i, \quad \text{where} \quad i,j=1,2,3 \tag{6}$$

The normal equations were generated by being expressed in matrix notation a

$$y = X\beta + \varepsilon \tag{7}$$

Then the vector of least squares estimators, **b**, that minimizes the errors was found as follows,

$$L = y'y - 2\beta' X' y + \beta' X' X\beta \tag{8}$$

Which was minimized by taking derivative of

$$\frac{\partial L}{\partial \beta}\Big| b = -2X'y + 2X'Xb = 0 \tag{9}$$

And $\hat{\beta}$ was estimated by

$$\hat{\beta} = (X'X)^{-1} X'y \tag{10}$$

Which result to the fitted regression model expressed as;

$$\hat{y} = b_0 + \sum_{j=1}^{k} b_i x_{ij}, \qquad i=1,2,......m \tag{11}$$

The difference between the observation y_i and the fitted value $\hat{y}_i$ results to residual, which is denoted as

$$\varepsilon_i = y_i - \hat{y}_i \tag{12}$$

Validation of the first order model was done by use of R^2 and Adjusted R^2 which used the formulas below;

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SEE}{SST} \tag{13}$$

And adjusted R^2 Statistics= $\dfrac{1 - SEE/(n-P)}{SST/(n-1)}$ (14)

And the predicted variance of the model was hence determined by

$$\text{var}[\hat{y}(x)] = \sigma^2 f'(x)(X'X)^{-1} f(x) \tag{15}$$

Analysis of variance (ANOVA) was used to check the adequacy of the model for the responses in the experimentation and was worked out using method of decomposition which was expressed as follows;

$$\left.\begin{aligned}
&SSE = SST - SSR \\
&SST = Y'Y - \frac{\left(\sum_{i=1}^{n} Y_i\right)^2}{n} \\
&SSR = \hat{\beta} X'Y - \frac{\left(\sum_{i=1}^{n} Y_i^2\right)}{n} \\
&Then \;\; \text{SSE}=Y'Y - \frac{\left(\sum_{i=1}^{n} Y_i\right)}{n} - \hat{\beta} X'Y - \frac{\left(\sum_{i=1}^{n} Y_i^2\right)}{n} \\
&SSE = Y'Y - \hat{\beta} X'Y
\end{aligned}\right\} \qquad \textbf{(16)}$$

Least significant difference (LSD) was used to separate data into significantly different groupings. Quartile plots were used to present data for this design (Khuri etal., 2010). Hypothesis testing for $H0$: β1 - β2 = ⋯ = βk =0 was compute as follows;

$$F_0 = \frac{SSR/K}{SSE/(N-K-1)} = \frac{MSR}{MSE} \qquad (17)$$

This test is for one factor at time, hence t-test was the most appropriate, *which* was expressed as

$$t_0 = \frac{\hat{\beta}_i}{\sqrt{\hat{\sigma}^2 W_{ii}}} \qquad (18)$$

3.4 Steepest Ascent

To assess the direction of response, the steepest ascent adopted and was derived from the fitted a first order model

$$\hat{y} = \hat{\beta}_0 + \sum_{i=1}^{k} \hat{\beta}_i x_i \qquad (19)$$

This was followed by use of the contours to present the direction of steepest ascent in which the response y increases most rapidly. The method of steepest assent for this study was derived as follows

$$Max \;\; \text{G}=\beta_0 + \sum_{i=1}^{k} \hat{\beta}_i x_i - \lambda[\sum_{i=1}^{k} x^2{}_i - r^2] \qquad (20)$$

Where λ is a LaGrange multiplier. Taking the derivatives of G yields $x_i = \frac{\hat{\beta}_i}{2\lambda}$

3.5G-optimal criteria

To evaluate the optimal criteria for the first order model, the D-optimality criterion was used to maximizes the determinant of the matrix X'X . The optimality of D-optimal design was measured by D-efficiency which was given by

$$DE = 100\left(\frac{|X'X|}{N}\right)^{1/p} \tag{21}$$

4.0 Results

4.1 Establishment of the effects of milk fat content, number of rotations and time on growth of kefir grains

The variables and data for this experiment was displayed in the table below

Table 4.1: Experimental variables and data on growth of kefir grains

Run	F	R	T	Yield	%Growth
1	3.5	50	36	48.33	142%
2	3.5	150	36	67.7	239%
3	10.5	50	36	66.33	232%
4	10.5	150	36	63.65	218%
5	3.5	100	24	64.09	220%
6	3.5	100	48	55.96	180%
7	10.5	100	24	82.7	314%
8	10.5	100	48	64.33	222%
9	7	50	24	68	240%
10	7	50	48	59.7	199%
11	7	150	24	84	320%
12	7	150	48	62	210%
13	7	100	36	65.67	228%
14	7	100	36	65.67	228%
15	7	100	36	65.67	228%

From table 4.1, the variables were denoted using F, R, and T to represent fats, rotations and incubation time respectively. The results from this table indicates that the highest growth rate was 320% which yielded 84 grams of kefir grains, using 7 grams milk fat content, speed of 150 rotations per minute and it took 24 hours as the incubation time. The growth rate was ;

$$\left.\begin{array}{l} Run1 \\ \%G = \dfrac{48.33-20}{20} = 142\% \\ Run2 \\ \%G = \dfrac{67.7-20}{20} = 239\% \\ \mathrm{M} \\ Run15 \\ \%G = \dfrac{65.67-20}{20} = 228\% \end{array}\right\}$$

Where 20 grams was the initial weight of the kefir grains before growth.

4.2 Coded variables and data for fitting first-order model

The coded variable were used to fit the first order model was displayed in the table below.

Table 4.2 Process Data for fitting the First-Order Model

Natural variables				Coded Variables			
Run	F	R	T	A	B	C	Yield
1	3.5	50	36	-1	-1	0	48.33
2	3.5	150	36	-1	1	0	67.70
3	10.5	50	36	1	-1	0	66.33
4	10.5	150	36	1	1	0	63.65
5	3.5	100	24	-1	0	-1	64.09
6	3.5	100	48	-1	0	1	55.96
7	10.5	100	24	1	0	-1	82.70
8	10.5	100	48	1	0	1	64.33
9	7.0	50	24	0	-1	-1	68.00
10	7.0	50	48	0	-1	1	59.70
11	7.0	150	24	0	1	-1	84.00
12	7.0	150	48	0	1	1	62.00
13	7.0	100	36	0	0	0	65.67
14	7.0	100	36	0	0	0	65.67
15	7.0	100	36	0	0	0	65.67

From table 4.2,the natural variables were converted in to coded variables in order to simplify the calculation, it was appropriate to use coded variables for describing independent variables in the (-1, 0, 1) interval. These were rescaled by taking 0 is in the middle of the center of the design, and ±1 as the distance from the center with direction. The transformation of F,R, and T variables to coded variable were obtained by

A= $\dfrac{F-7}{3.5}$; B= $\dfrac{R-100}{50}$; and C= $\dfrac{T-36}{12}$

And complete calculation of the coded variables was achieved as follows;

$$\left.\begin{array}{l}
Run\ 1 \\
A=\dfrac{3.5-7}{3.5}=-1;\quad \text{B}=\dfrac{50-100}{50}=-1;\quad \text{and}\ \ \text{C}=\dfrac{36-36}{12}=0 \\
Run\ 2 \\
A=\dfrac{3.5-7}{3.5}=-1;\quad \text{B}=\dfrac{150-100}{50}=1;\quad \text{and}\ \ \text{C}=\dfrac{36-36}{12}=0 \\
\vdots \\
Run\ 15 \\
A=\dfrac{7-7}{3.5}=0;\quad \text{B}=\dfrac{100-100}{50}=0;\quad \text{and}\ \ \text{C}=\dfrac{36-36}{12}=0
\end{array}\right\}$$

The first run indicates that the experiment was performed using milk of low fat content (-1), low speed of rotation (-1) and medium incubation time (0).The process continues as the growth is being recorded. From the results of table 4.2 it indicates that the highest growth was recorded at medium milk fat content(0),high rotation speed(1) and low incubation time(-1).This yielded a growth of 84 grams of kefir grains.

4.3 First-Order Design to study effects of culture conditions

To determine the effect of milk fat content ,number of rotations and incubation time ,the first –order model was designed so that the effects of the culture conditions on the growth of kefir grains using milk was determined . The fitting of the first –order model was calculated and the results were as follows;

$$\text{X}=\begin{vmatrix} 1 & -1 & -1 & 0 \\ 1 & 1 & -1 & 0 \\ 1 & -1 & 1 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & -1 & 0 & -1 \\ 1 & 1 & 0 & 1 \\ 1 & -1 & 0 & -1 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & -1 & -1 \\ 1 & 0 & -1 & 1 \\ 1 & 0 & 1 & -1 \\ 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{vmatrix} \quad \text{and} \quad \text{Y}=\begin{vmatrix} 48.33 \\ 67.70 \\ 66.33 \\ 63.65 \\ 64.09 \\ 55.96 \\ 82.70 \\ 64.33 \\ 68 \\ 59.7 \\ 84 \\ 62 \\ 65.67 \\ 65.67 \\ 65.67 \end{vmatrix}$$

$$\text{X}'\text{X}=\begin{vmatrix} 15 & 0 & 0 & 0 \\ 0 & 8 & 0 & 0 \\ 0 & 0 & 8 & 0 \\ 0 & 0 & 0 & 8 \end{vmatrix} \quad X'Y=\begin{vmatrix} 983.80 \\ 40.93 \\ 34.99 \\ -56.80 \end{vmatrix} \quad \text{and} \ (X'X)^{-1}=\begin{vmatrix} 0.0667 & .0000 & 0.0000 & 0.0000 \\ 0.0000 & 0.1250 & 0.0000 & 0.0000 \\ 0.0000 & 0.0000 & 0.1250 & 0.0000 \\ 0.0000 & 0.0000 & 0.0000 & 0.1250 \end{vmatrix}$$

Thus $\hat{\beta} = \begin{vmatrix} 65.58 \\ 5.12 \\ 4.37 \\ -7.10 \end{vmatrix}$

The fitted regression model for growth yield was fitted as

Y= 65.58 + 5.12 A + 4.37B -7.10C

Where Y is the growth of kefir grain

A is the coded variable for fat content in milk

B is the coded variable of number of rotations of the culture experiments

C is the coded variable of time of reactions for the experiments.

From the model, milk fat content and speed of rotations have positive effect while incubation time has a negative effect on the growth of kefir grains using milk. This indicates that the growth of kefir grains increases steadily with increase in fat content and speed of rotations but decreases steadily with the increase in incubation time after some period of time for growing kefir grains.

4.3.1 Model Adequacy Checking

Adequacy of the model was analyzed to examine the fitted model if it provides an adequate approximation of the true response surface (the growth of kefir grains). The normality used to process the results was, analysis of variance, regression analysis, and lack of fit test that examined on (4.5). R-program was used to conduct the regression analysis and the variance of analysis of growth of kefir grain. The results were displayed in the table below.

Table 4.3: R – program output on regression analysis to check model adequacy

Call:

lm (formula = Y ~ A + B + C)

Residuals:

Min	1Q	Median	3Q	Max
-11.4267	-0.5867	0.0833	2.7227	6.9396

Coefficients:

<table>
<tr><th></th><th>Estimate</th><th>Std. Error</th><th>t value</th><th>Pr(>|t|)</th><th></th></tr>
<tr><td>(Intercept)</td><td>65.587</td><td>1.398</td><td>46.905</td><td>5.07e-14</td><td>***</td></tr>
<tr><td>A</td><td>5.116</td><td>1.915</td><td>2.672</td><td>0.02171</td><td>*</td></tr>
<tr><td>B</td><td>4.374</td><td>1.915</td><td>2.284</td><td>0.04321</td><td>*</td></tr>
<tr><td>C</td><td>-7.100</td><td>1.915</td><td>-3.708</td><td>0.00345</td><td>**</td></tr>
</table>

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.416 on 11 degrees of freedom Multiple R-squared: 0.7036, Adjusted R-squared: 0.6227 F-statistic: 8.703 on 3 and 11 DF, p-value: 0.003043

From table 4.3, the intercept of the first order model is very significant at p-value<0.001, milk fat content ,and speed of rotation are significant at p-value<.05 and incubation time effect is significant at p-value<.01.

The components in a nova table were calculated using the method of decomposition of variance for the kefir growth as;

SST=1088.34; SSR=765.73; SSE=322.61; MSR = 255.243 and MSE = 29.328

Therefore, using (3.22) the statistic $F\ calculated$ =8.703

The results were then displayed in the a nova table as in table below

Table 4.4 Analysis of Variance for Significance of Regression

Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0
Regression	*255.243*	*3*	*255.243*	*8.703*
Error or Residuals	*322.61*	*11*	*29.328*	
Total	*577.853*	*14*		

From table 4.4, the calculated value of the F-ratio of the regression model is more than the standard value specified(F-table) at 95% confidence level, and thus the first order model was consider adequate.

4.3.2 The test for significance of regression

This was achieved by carrying out residual analysis and displaying the normal probability of residuals graph. For residual analysis, the residuals from a 3^3 design for three factors A, B, and C was obtained by estimating the regression coefficients β_A, β_B, and β_C by one-half the corresponding effect estimates, and β_0 being the grand average and expressed as ;

$$\hat{y} = 65.587 + \left(\frac{5.116}{2}\right)X_A + \left(\frac{4.374}{2}\right)X_B - \left(\frac{7.10}{2}\right)X_C$$

=65.587+2.558 X_A +2.187 X_B -3.55 X_C

For first run in table 4.2, we have A is at low (-1) level, B is at low (-1) and C is at medium (0), we substitute it in the above equation with the coded values for the first run then it yields

=65.587+2.558(-1) +2.187(-1)-3.55(0)=60.842 grams

And the observed yield was 48.33 hence using (3.16) the yield is -12.512 residuals for first run and the other 14 runs were similarly obtained and results were obtained as shown in the table below

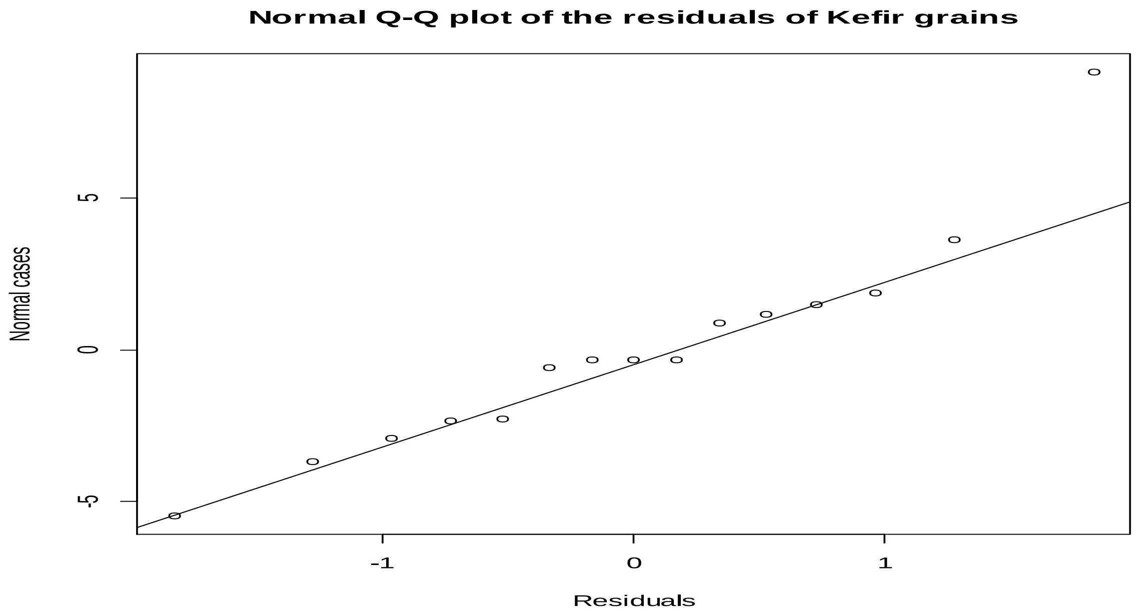
Table 4.5: Residuals of the experiments

Run	A	B	C	Yield	Pyield	Residual
1	-1	-1	0	48.33	60.842	-12.512
2	-1	1	0	67.7	65.216	2.484
3	1	-1	0	66.33	65.958	0.372
4	1	1	0	63.65	70.332	-6.682
5	-1	0	-1	64.09	66.579	-2.489
6	-1	0	1	55.96	59.479	-3.519
7	1	0	-1	82.7	71.695	11.005
8	1	0	1	64.33	64.595	-0.265
9	0	-1	-1	68	66.95	1.05
10	0	-1	1	59.7	59.85	-0.15
11	0	1	-1	84	71.324	12.676
12	0	1	1	62	64.224	-2.224
13	0	0	0	65.67	65.587	0.083
14	0	0	0	65.67	65.587	0.083
15	0	0	0	65.67	65.587	0.083

From table 4.5, residuals for stationery points i.e. run 13, 14, and 15 had the lowest residuals of .083, hence they gave the best results for the predictive response that was not far from the observed responses. This indicates that for predictive response then the experiment should be carried out at medium levels for all the factors. In general the range of the $|Residuals|$ is (12.512-0.083)=12.429 which indicates that the growth of kefir grains using milk in this experiment was in good agreement with the predicted ones, suggesting a satisfactory and accurate first –order model for this study.

The normal probability graph for test of the significant of the first order model was displayed in figure below.

Figure 4.1 Normal graph of the residual of growth of kefir grains



From figure 4.1, the normality assumption is tested such that if the residuals plot approximately along a straight line, then the normality assumption is satisfied. In this study, the residuals can be judged as normally distributed; therefore normality assumptions for both of the responses are satisfied. Therefore, the test for the significance of the regression can be applied to determine if the relationship between *y* and $x_1, x_2....., x_q$ exists.

As a result, the hypothesis for the statistical analysis of response variable for this study was stated as:

$H_0 : \beta_1 = \beta_2 = \beta_3 = o$ vs $H_1 : \beta_j \neq 0$ for at least one j.

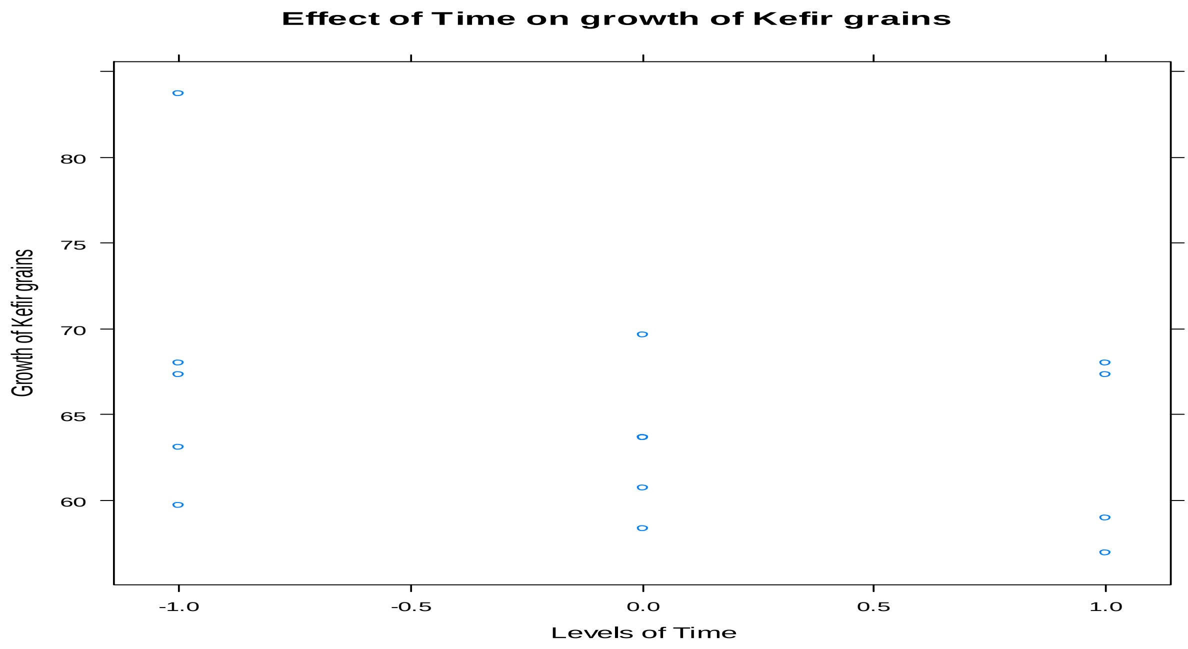
At the significant level α = 0.05, the critical value F (.05, 3, 11) = 3.59 < the calculated F = 8.703. Also, P-value from table 4.3 for the statistic F < 0.05. There is a significant statistical evidence to reject the null hypothesis. It implies that at least one of the independent variables – fat content (A), number of rotations (B), and time (C), contributes significantly to the model compared to other independent variables.

How well the estimated model fits the data was measured by the value of R^2. When R^2 is closer to 1, the better the estimation of regression equation fits the sample data. For the first order linear equation, R^2 = .703 and Adjusted - R^2 =.6227.These results are the same with the one in R output in table 4.3 above. Both of $R^2$and adjusted R^2 are statistically significant at p-value<.05 for the growth of kefir grains yield as indicated in table 4.3. R^2- adjusted R^2 = .08 is very small and this indicates that the regression equation fits the data very well.

4.3.3 The Test for Individual Regression Coefficients

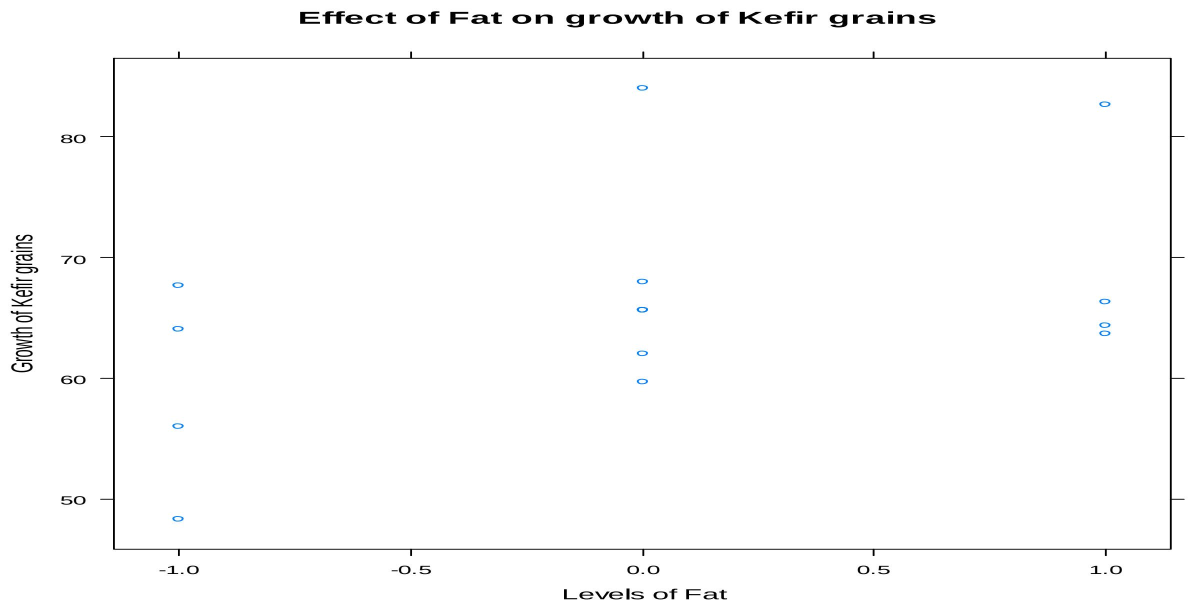
In order to determine whether given variables should be included or excluded from the first order model, a test hypothesis for the individual regression coefficients was performed. The simple analysis started with a main effects plot. The main effect was calculated by subtracting the overall mean for the factor from the mean for each level. The effects of each independent variable were displayed using the figures below.

Figure 4.2: Effects of Time on the growth of kefir grain



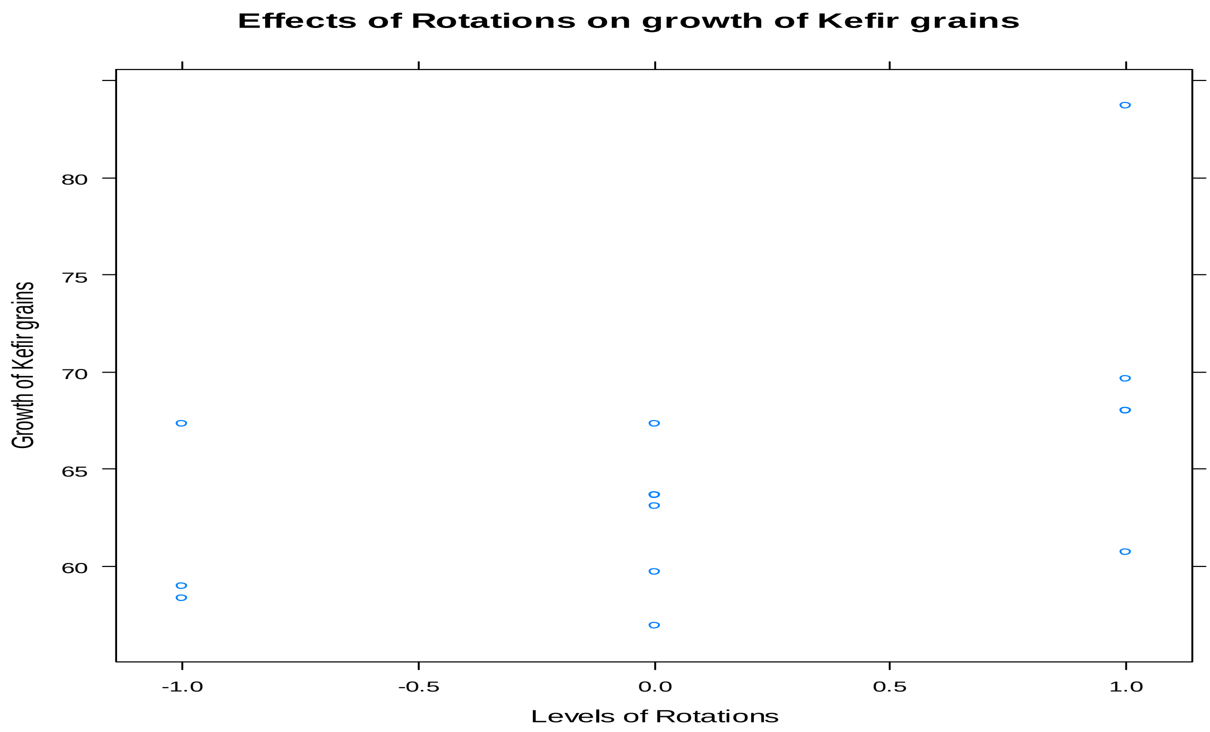
From figure 4.2, it shows that the grown of kefir was high at low time (-1) and low growth was experienced at a high incubation time (1).This indicates a negative effect of incubation time on growth of kefir grains.

Figure 4.3: Effects of fat on the growth of kefir grain



From figure 4.3, shows that the grown of kefir was high at medium(0) and high(1) levels of milk fat content and low growth at low(-1) level of milk fat content. This indicates a positive effect of milk fat content on growth of kefir grains.

Figure 4.4: Effects of Number of Rotations on the growth of kefir grain



From figure 4.4, shows that the grown of kefir was high at high (1) levels of rotations per minute and low growth at low (-1) level of rotations per minute. This indicates a positive effect of rotations per minute on growth of kefir grains.

In general the analysis indicates that the factors fat content (A) and number of rotations (B) increase when they move from the low level to the high level of growth of kefir grains .Each level of the factors affects the growth differently. Each factor at their high level results in higher growth compared to that at the low level. If the slope is close to zero, the magnitude of the main effect would be small and for this study it indicates a large effect since the coefficients are not close to zero. In order to determine the significance of the factors, a *t*-test, was conducted to identify the significance of the main factors since the hypotheses test was particular to one coefficient at a time. To examine the significant contribution of the independent variables to growth of kefir grains , the following calculations for the following hypotheses were performed:

$$\left.\begin{array}{ll} H_0 : \beta_{fats} = 0 & H_1 : \beta_{fats} \neq 0 \\ H_0 : \beta_{rotations} = 0 & H_1 : \beta_{rotations} \neq 0 \\ H_0 : \beta_{time} = 0 & H_1 : \beta_{time} \neq 0 \end{array}\right\}$$

The values of $se(\hat{\beta}_i)$ were found as follows . Recall from earlier calculation that $\hat{\beta}_i = \begin{vmatrix} 65.58 \\ 5.12 \\ 4.37 \\ -7.10 \end{vmatrix}$ and

$$\left(X\ 'X\right)^{-1} = \begin{vmatrix} 0.0667 & .0000 & 0.0000 & 0.0000 \\ 0.0000 & 0.1250 & 0.0000 & 0.0000 \\ 0.0000 & 0.0000 & 0.1250 & 0.0000 \\ 0.0000 & 0.0000 & 0.0000 & 0.1250 \end{vmatrix} \text{ and } \quad \hat{\sigma}^2 = 29.328$$

Since MSE= $\hat{\sigma}^2$

t-statistics were obtained as follows

$$t_A = \frac{5.12}{\sqrt{29.328*.0667}} = 3.6607$$

$$t_B = \frac{4.37}{\sqrt{29.328*.1250}} = 2.282$$

$$t_C = \frac{-7.10}{\sqrt{29.328*.1250}} = -3.708$$

These *t*-statistic values were compared with the *critical t*-values and the null hypothesis $H_0 : ß_j = 0$ was rejected if the observed $|t0|$ > critical value $t_{\frac{\alpha}{2}, N-q-1}$, α = .05 from t-test table.

$|t_A|$ =3.6607> $t_{.025,11}$= 2.201 and

$|t_B|$ = 2.282> $t_{.025,11}$ =2.201 and

$|t_C|$ = 3.708> $_{.025,11}$ = 2.201 and

The null hypotheses H_0: $ß_{fat} = 0$,H_0: $ß_{rotations} = 0$ and H_0: H_0: $ß_{time} = 0$, were all rejected since their observed $|t0|$ > critical value $t_{\frac{\alpha}{2}, N-q-1}$. It was then concluded that the independent variables: fat (*A*), and Rotations (*B*) and time(C) contribute significantly to the response growth of kefir grains using milk.

Thus the best model for first-order model for the main effects is

Y= 65.58 + 5.12 A + 4.37B -7.10C and Y yield= 65.58 + 5.12 (1) + 4.37(1) -7.10(-1)

=82.17 grams.

4.3.4 The Main Effects and Interactions for 3^3 Design

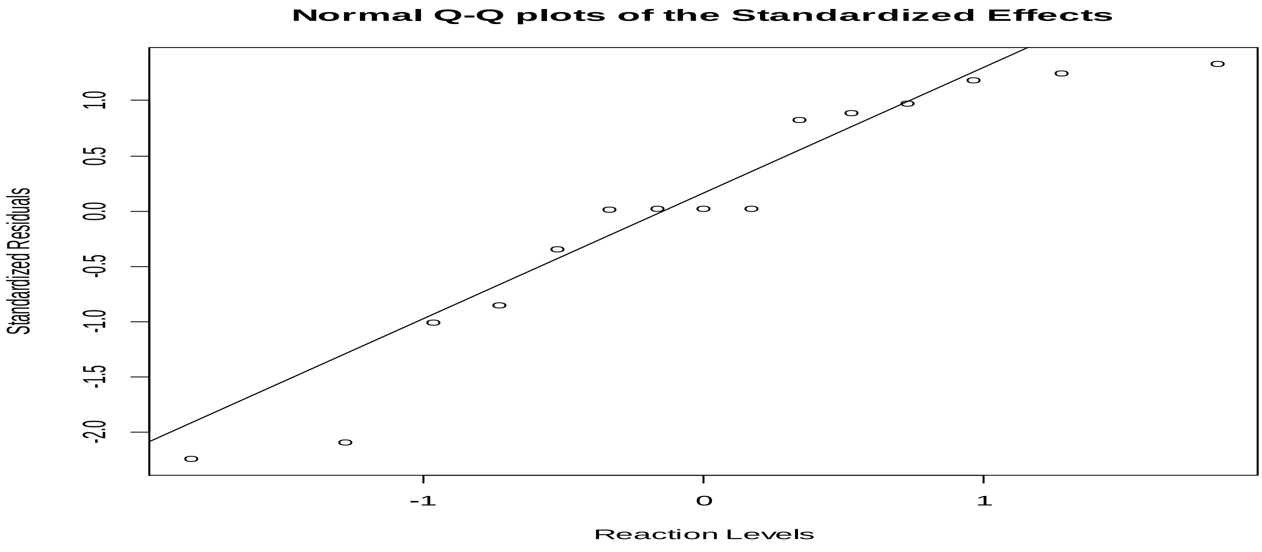
The data was analyzed to assess the effects of main and interaction of variables. The data of main effects and their interactions were displayed in Table below.

Table 4 .6 Data of the Main Effects and Interactions for 3^3 Design

Run	A	B	C	AB	AC	BC	ABC	Yield
1	-1	-1	0	1	0	0	0	48.33
2	-1	1	0	-1	0	0	0	67.70
3	1	-1	0	-1	0	0	0	66.33
4	1	1	0	1	0	0	0	63.65
5	-1	0	-1	0	1	0	0	64.09
6	-1	0	1	0	-1	0	0	55.96
7	1	0	-1	0	-1	0	0	82.70
8	1	0	1	0	1	0	0	64.33
9	0	-1	-1	0	0	1	0	68.00
10	0	-1	1	0	0	-1	0	59.70
11	0	1	-1	0	0	-1	0	84.00
12	0	1	1	0	0	1	0	62.00
13	0	0	0	0	0	0	0	65.67
14	0	0	0	0	0	0	0	65.67
15	0	0	0	0	0	0	0	65.67

Analysis of the main and interaction effects for 3^3 Design was performed by examining the normality of the estimated effects. This was done by use of a normal Q-Q plot to determine the statistical significance of both main and interaction effects. The effects that were found not significant were falling far away from the line. The Figure 4.5 illustrates the normal Q-Q plot of these effects.

Figure 4.5: The normal Q-Q plot of the effects of variables on growth of kefir grains



From this analysis it was concluded that the main effects *A*, *B*, and *C*, the interactions *AB*, *BC*, and *BC* are significant. Since majority of the points lie close to the line, therefore their contribution had an effect on the first order- model that was used to predict the growth of kefir grains using milk and fat, rotations and time as the control factors.

4.3.5: Analysis of significance of data using the sparsity of effects principle

For this study, the assumption was that the highest interaction component *ABC* effect was negligible and its mean square can be used to obtain an error term. The results was processed using R and were displayed in table 4.7 and table 4.8 as shown below

Table 4.7: Regression estimates of main and interaction effects

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	65.587	1.032	63.524	4.19e-12 ***
A	5.116	1.414	3.619	0.00679 **
B	4.374	1.414	3.094	0.01481 *
C	-7.100	1.414	-5.022	0.00102 **
AB	-5.513	1.999	-2.757	0.02479 *
AC	-2.560	1.999	-1.280	0.23628
BC	-3.425	1.999	-1.713	0.12506
ABC	NA	NA	NA	NA

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.999 on 8 degrees of freedom Multiple R-squared: 0.8825, Adjusted R-squared: 0.7943 F-statistic: 10.01 on 6 and 8 DF, p-value: 0.002351

The results from table 4.7 yielded were fitted as follows;

$$Y=65.587+5.116A+4.374B-7.100C-5.513AB-2.560AC-3.425BC$$

Where the interaction of ABC has no influence on the growth of kefir grain using milk as the culture liquid. The values for R-squared: 0.8825, Adjusted R-squared: 0.7943 were high, i.e. very close to 1 and hence indicates that the model generated fits the data very well. The significant of the effects of the factors were evaluated using anova and the results were displayed as in the table below.

Table 4.8 Analysis of Variance Table

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
A	1	209.41	209.41	13.0962	0.006794	**
B	1	153.04	153.04	9.5708	0.014807	*
C	1	403.28	403.28	25.2208	0.001024	**
AB	1	121.55	121.55	7.6017	0.024785	*
AC	1	26.21	26.21	1.6394	0.236278	
BC	1	46.92	46.92	2.9345	0.125059	
Residuals	8	127.92	15.99			

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

As it appeared from a nova table and regression estimates tables, the interaction between milk fat content and number of rotations per minute is significant at 5% level of significance, while interaction between AC and BC are not significant at p-value<.05. Meanwhile, the 3- factor interactions do not even appear to contribute significantly to the model although it has no effect at all. The *F*-tests reveal that main effects are significant at p<.05. The following estimated equation was thus the final model for the main and interaction effects of the model for the response variable Yield of kefir grains

Y=65.587+5.116A+4.374B-7.100C-5.513AB and

Yield=65.587+5.116(1)+4.374(1)-7.100(-1)-5.513(1)(1)

=76.664 grams

4.3.6Determination of direction of growth for kefir grains using milk

This was enabled by computing the steepest ascent for the growth. The sign of steepest descent requires the direction to be opposite of the sign of the coefficient. Thus for the first order model $\hat{y} = 65.59 + 5.12A + 4.37B - 7.10C$,the path of steepest ascent results in $\hat{y}$ moving in a positive direction for increase in milk fat content (A)and number of rotations(B)and for a decrease in time(C).Also $\hat{y}$ (predicted growth in kefir grain) increases 5.12 times as fast as Δ increase in milk fat content,4.57 times as fast as Δ increase in number of rotations and -7.1 times as fast as Δ decrease in time duration for the incubation process. This follows that $\hat{y} = 65.59 + 5.12A + 4.37B - 7.10C$ has to be moving away from the design center, the point (*A* = 0, *B* = 0, *C*= 0), along the path of steepest ascent, we would move -7.1 units in the time (C) direction for every 4.37 units in the milk fat content (*A*)direction and 5.12 units in the number of rotations(*B*)direction calculated as follows

$$\Delta_A = \frac{A}{C}\Delta_C = \frac{5.12}{-7.1} = -.72 \quad \text{and } \Delta_B = \frac{B}{C}\Delta_C = \frac{4.37}{-7.1} = -.62$$

The method of steepest ascent assumes that the experimenter wishes to move from the center of the initial design in the direction of the point that indicates the maximum predicted increase in the response and for steepest descent indicates that the maximum predicted decrease in the response.Therefore the path moves 5.12 units in A every -7.1 units in C direction and moves 4.37 units in B direction for every -7.1 units in C direction away from the origin(0,0,0).Or ,rescaled,-.72 units in the A direction for every 1 unit in C direction and -.62 units in the B direction for every 1 unit in C direction from the origin(0,0,0).

To determine the path we first find the coding values of the factors which was calculated as follows

$$\left.\begin{array}{l} x_A = \dfrac{\xi_1 - 7}{7}; \quad \xi_A = 7+7x_A \\ x_B = \dfrac{\xi_B - 100}{100}; \quad \xi_B = 100+100x_B \\ x_C = \dfrac{\xi_C - 36}{24}; \quad \xi_C = 36+24x_C \end{array}\right\}$$

Taking 36 in ξ_C as the basic step size in natural units, thus, $\Delta c = -1$ step size in coded units. Then $\xi_{A=.72}$ and $\xi_{B=.62}$ step sizes in coded units as follows;

$$\Delta_A = \frac{A}{C}\Delta_C = \frac{5.12}{-7.1}(-1) = .72 \quad \text{and } \Delta_B = \frac{B}{C}\Delta_C = \frac{4.37}{-7.1}(-1) = .62$$

In natural units the step sizes for milk fat content and number of rotation are $\Delta_A = 3.5$ and $\Delta_B = 50$

The results of steepest ascent were displayed as shown in the table below.

Table 4.9: Procedure along the Point of Steepest Ascent

PSA	x1	x2	x3	F(ξ_A)	R(ξ_B)	T(ξ_C)	RESPONSE(y)
(0,0,0)	0	0	0	7	100	36	61.0
(0,0,0)+ Δ	0.72	0.62	1	12.04	162	72	64.7
(0,0,0)+2 Δ	1.44	1.24	2	17.08	224	108	67.1
(0,0,0)+ 3 Δ	2.16	1.86	3	22.12	286	144	69.65
(0,0,0)+ 4 Δ	2.88	2.48	4	27.16	348	180	73.6
(0,0,0)+5 Δ	3.6	3.1	5	32.2	410	216	79.9
(0,0,0)+ 6 Δ	4.32	3.72	6	37.24	472	252	80.5
(0,0,0)+ 7 Δ	5.04	4.34	7	42.28	534	288	82.96
(0,0,0)+ 8 Δ	5.76	4.96	8	47.32	596	324	76.38
(0,0,0)+9 Δ	6.48	5.58	9	52.36	658	360	75.04

From table 4.9, move Δ along the PSA from (0, 0, 0).This is the point (0, 0, 0) + Δ = (0.72, .62, 1) which in original scale is (12.04 fat content, 162 rotation, 72 hours).The response is collected at (12.04 fat content, 100 rotation, 36 hours) was $y = 61$ grams. Next move Δ along the PSA from (0, 0, 0)+ Δ to (0, 0, 0)+ 2Δ the new point 2*(0.72,.62,1)=(1.44,1.24,2) which in the original scale is (17.08g, 224r, 108h)the response collected at that point is $y = 64.7$ grams ,which is larger than $y = 61$ grams. The trend continues to increase until the 8th point that is from (0, 0, 0)+ 7Δ to (0, 0, 0)+ 8Δ the first decrease is observed which is from (5.04,4.34,7) to (5.76,4.96,8) which in the natural scale is (42.28,534,288) to(47.32,596,324).The response decreases from 82.96 to 76.38.One more point along the PSA as further evidence for finding a new path at (0, 0, 0)+ 9Δ =(6.48,5.58,9) which the natural scale is (52.36 ,658,360) at response $y = 75.04$ which is smaller than 76.38.This lead to stopping collection of data along the path.

4.3.7 Optimal Criteria

D-optimality criterion was used to maximizes the determinant of

$$\frac{|X'X|}{N} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & .53 & 0 & 0 \\ 0 & 0 & .53 & 0 \\ 0 & 0 & 0 & .53 \end{pmatrix}$$

Thus D-optimal=1

Optimality of D-optimal design was measured by D-efficiency and results were obtain as

$$DE = 100(1)^{1/9} = 100\%$$

Hence the model achieved the standard fractional factorial of 100% efficiency

5.0 Conclusion

Response Surface Models evolved for responses show the effect of each input parameter and its interaction with other parameters, depicting the trend of response. Verification of the Fitness of each model using ANOVA technique, shows that all the models were significant at p-value <.05.This indicates that all the models were good for navigating within the design space of the experiment. Further validation of the models was done with the additional experimental data collected and it demonstrates that the models have high reliability for adoption within the chosen range of parameters since their adjusted R^2 were found to be close to 1 and hence they are good for prediction purpose. A first-order model was used to describe some part of the response surface during the growth of kefir grains and it fitted the data by least squares method. The regression estimates of the first order model indicated that milk fat content and speed of rotation had a positive significant effects and incubation time had a negative significant effect on the growth of kefir grains.

Acknowledgement

I would like to thank Almighty God who gave me strength, good health, and wisdom during period of my research. Thanks to my supervisor, Professors Koske J.K and John Muindi Mutiso, whose tolerance and help has been invaluable.

Reference

A. Lewis (2012). Optimization of cocoa beans roasting process using Response Surface Methodology based on concentration of pyrazine and acrylamide. *Malaysia*

Box GEP, Behnken DW(1960). "Some New Three Level Designs for the Study of Quantitative Variables." Technometrics, 2, 455-475.

Box GEP, Hunter WG, Hunter JS (2005). Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building. 2nd edition. John Wiley & Sons, New York.

Chigbu, P.E. and Nduka, U.C. (2006).On the Optimal choice of Cube and Star replications in restricted second-order designs. (preprint, IC/2006/113) The Abdul Salam International Centre for Theoretical Physics, December 2006.

Corthésy B, Gaskins HR, Mercenier A. Cross-talk between probiotic bacteria and the host immune system. J Nutr. 2007; 137(3 Suppl 2), 781S–90S. pmid:17311975

Dominic N Anfiteatro [dna] 1999-2015. How to make kefir and Recipes Created, published, maintained and Copyrighted © by. All rights reserved

Farnworth ER. Kefir–A complex probiotic. Food Sci Technol. Bulletin: Functional Foods. 2005; 2(1): 1–17. doi: 10.1616/1476-2137.13938

Khuri, A. I. and S. Mukhopadhyay (2010) Response surface methodology, *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(2), 128–149.

Kiefer, J. (1974). General equivalence theory for optimum designs (approximate theory). Annals of Statistics, 2:849-879.

Kiefer, J. and Wolfowitz, J. (1959). Optimum designs in regression problems. Ann. Math. Statist.,30:271{294.

Joe Leech(,2015).Health Benefits of Kefir

Molina S, Moran-Valero M I, Martin D, Vázquez L, Vargas T, Torres C F et al. Antiproliferative effect of alkylglycerols as vehicles of butyric acid on colon cancer cells. ChemPhys Lipids. 2013; 175–176: 50–56. doi: 10.1016/j.chemphyslip.2013.07.011. pmid:23973778

M. Cavazzuti, *Optimization Methods: From Theory to Design*, 13 DOI: 10.1007/978-3-642-31187-1_2, © Springer-Verlag Berlin Heidelberg 2013

Montgomery, D.C. (2005). Design and analysis of experiments (6th ed.). Massachusetts: John Wiley.

Pengpeng qiu (2014)Application of Box–Behnken design with response surface methodology for modeling and optimizing ultrasonic oxidation of arsenite with H2O2,Cent.Europe journal of chemistry.

Palanikumar.K(2008). Application of Taguchi and response surface methodologies for surface roughness in machining glass fiber reinforced plastics by PCD tooling, International Journal of Advanced Manufacturing Technology, 36, 19–27

Rodriguez, M., Jones, B., Borror, C.M., and Montgomery, D.C. (2010), "Generating and Assessing Exact G-Optimal Designs", Journal of Quality Technology, Vol. 42, No. 1, pp. 329

Tanyildizi MS, Ozer D, Elibol M (2005) Optimization of -amylase production by Bacillus sp. using response surface methodology. Process biochemistry 40: 2291-2296.

Van Baarlen, P.*et al*. Human mucosal *in vivo* transcriptome responses to three lactobacilli indicate how probiotics may modulate human cellular pathways. Proc. Natl Acad. Sci. USA108 (Suppl. 1), 4562–4569 (2011).

Victorbabu and Ch. V. V. S. Surekha(2013)a note on measure of rotatability for second order response surface designs using incomplete block designs

Venter G (1998) Non-dimensional response surfaces for structural optimization with uncertainty. PhD Thesis, University of Florida

V. N. Gaitonde & S. R. Karnik & B. Siddeswarappa, B. T. Achyutha, Integrating Box-Behnken design with genetic algorithm to determine the optimal parametric combination for minimizing burr size in drilling of AISI 316L stainless steel, Int Journal of Advanced Manufacturing Technology 37(2008),230–240

Welman, A. D., & Maddox, I. S. (2003). Exopolysaccharides from lactic acid bacteria:Perspectives and challenges. Trends in Biotechnology, 21(6), 269–274.

Appendix

APPENDIX 1: Size and Samples of Kefir grains

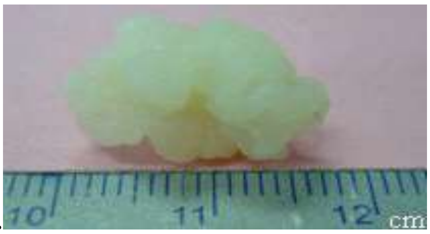
.

Fig 1.1: Size of kefir grain