DETERMINATION OF THE MOST IMPORTANT FACTORS AFFECTING DAIRY PRODUCTS DURING THE STORAGE PERIOD USING FACTORIAL EXPERIMENTS: CASE STUDY IN IRAQ

Bahr Kadhim MOHAMMED [1]
Meshal Harbi ODAH [2]
Ali Sadig Mohommed BAGER [3]

ABSTRACT
Dairy products are at the forefront of food that is susceptible to rapid damage. It is suitable for bacterial growth. Any defect in the validity, storage method or safety of the methods used in all stages of manufacturing causes product corruption immediately. Product storage is indispensable in many factories. This paper deals the most important factors that affect the damage of dairy products during storage (before marketing). The factors are described as follows (temperature, humidity, light intensity, package size, and storage period). Each factor has two levels (high, low). In such studies it is necessary to determine the effect of the main factors and their interactions on the damage of the product. Models factorial experiments are consider of the important statistical methods to study the main and interactions effects together, especially when studying the levels of these factors, so we use in this paper a model of factorial experiment of type(2^5), to analyze this type of experiments uses the technique multiple regression model and estimation of the parameters by the of ordinary least squares method (OLS). The results of the analysis showed that there are two factors had a main effect on product damage, as well as five bilateral interactions and five triple interactions that had a clear effect on the dairy product damage.

KEYWORDS: *factorial experiments, dairy products, multiple regression model, main effects and interactions, ordinary least square.*

JEL CLASSIFICATION: *C13, C30 ,C90, M11*

1. INTRODUCTION

Product storage is indispensable in many factories. These factories retain their product from stock until they are marketed. The process of preserving the stored materials from damage and ensure the safety and survival of the goods from the time they enter the warehouse until it is disbursed to the requesting parties is also very important. However, inventory control may not be appropriate in some companies, despite the impact on the product. Many products are damaged after being manufactured during storage, including dairy products.
There are many statistical methods that are used to study and determine the effect of factors or variables on the response variable. But it does not take into account the joint effects of these factors on the studied phenomenon. In 1926, Fisher used the factorial experiments models to study the effect of common factors, and concluded that it were more efficient than studying the effect of one factor at a time. Yates (1937) who has great credit in promoting the development and analysis of the factorial experiments, used the method of statistical analysis of the factorial experiments of a wide

[1] The Bucharest University of Economic Studies, University of AL-Qadisiyah, Iraq, baherm@yahoo.com.
[2] The Bucharest University of Economic Studies, Muthanna University, Iraq, m.algelidh@gmail.com.
[3] The Bucharest University of Economic Studies, Muthanna University, Iraq, nader.ali62@yahoo.com.

range of the type 2^2, 3^2. The researchers then continued to develop and employ factorial experiment models in many fields including agricultural, medical, industrial and other.

In this paper, we use factorial experiments models to study the effect of the five main factors with their interactions with two level for each factor on the damage of the dairy product during the storage period. Thus, we must choose a model for this experiment of type (2^5). In this type of factorial design we have five factors each at two levels, so that there are $2 \times 2 \times 2 \times 2 \times 2 = 32$ treatment combinations. Thus, to analyze this experiment, we must add four replicated runs (the last four runs) at the center (0). The addition of replicated center points allows an independent estimate of error to be obtained without affecting the estimates of the factorial effects. Generally, three to five center runs are recommended (Montgomery, 2009). Using these center points, we can obtain an estimate of the variability, analysis of variance and conduct a lack-of-fit test to these experiments.

This paper is organized as follows. In the second section, we explain briefly the concept of factorial experiment and some types of factorial experiments, such as full factorial design. In Section 3 we provide regression experiments design and the relationship between experiment of design with regression model, as well as estimating parameters. The fourth section includes the study sample with a description of the factors and their interactions. In the fifth section we present a data statistical analysis and summarize the results of a sample study. A brief conclusion is included in sixth section.

2. FACTORIAL EXPERIMENTS DESIGN

Factorial experiments design represents the introduction of more than one factor in the experiment, and since every factor consists of a number of levels, the processors consist of possible combinations between (Levels) of these (Factors). Factorial experiences are obtained through a factorial experiment on the effects of the (interactions) and these interactions have great importance for the interpretation of the data in the experiment which cannot be neglected or ignored. Factorial design also provides a systematic method of investigating the relationships among the effects of different factors (i.e., interactions) (Wu & Hamada, 2009). The effect of a factor can be defined as the change in response produced by a change in the level of the factor. This is referred to as the main effect. In some experiments, it may be found that the difference in the response between levels of one factor is not the same at all levels of the other factors. This is referred to as an interaction effect between factors. Collectively, main effects and interaction effects are called the factorial effects (Wu & Hamada, 2009). A factorial design can estimate all main effects and interactions to these factors, therefore, factorial experiments are conducted in a wide range of experimental designs, which regulate the appropriate manner for simple designs. The factorial experiments are obtained through interactions between factors and levels. These interactions have great importance for the interpretation of the data in the experiment and cannot be neglected or ignored, since they are used in order to find the impact that results from the variables participation at the impact on the studied phenomenon (Douglas, 2001).There are many types of factorial experiments, depending on the number of factors and the number of levels.

In factorial design with k factors, a common and reasonable assumption is that higher-order interactions are assumed to be negligible due to the fact that they are less likely to be important than lower-order interactions (Wu & Hamada, 2009). Thus in this study we can estimate all main effects and all two-factor interactions and three-factor interactions assuming that fourth-order interactions and five-order interactions are negligible, which is quite reasonable in practice. The factorial design assumes that the dependent variable (responses) approximates the normal distribution. This case can be verified by checking graphically (either a histogram with normal distribution curve, or with a Q-Q-Plot) or it can be tested with a goodness of fit test against normal distribution such as Kolmogorov-Smirnov test (Massey & Frank, 1956).

3. REGRESSION EXPERIMENT DESIGN

The classical methods use the approach of multiple regression to define the concept of main effects and interaction effects for two-level designs (Mason et. al., 2003; Williams, 1970). The analysis of experiments can be formulated in the form of the general linear regression where we have n observation on a dependent variable Y and predictors ($X_1, X_2, \dots, X_P$), and

$$Y = X\beta + e \qquad (1)$$

where $e \sim N(0, \sigma^2)$ and $\beta = (\beta_1, \beta_2, \dots, \beta_p)$. Throughout this work, we center each input variable so that the observed mean is zero and standard deviation is one.

In this paper, we use factorial design consists of five factors symbolized by 2^5. The factors expressed in capital letters A, B, C, D, and E with two levels for each factor. There are 2^5 =32 treatments or level combinations. A common regression model for studying main effects and interactions is:

$$y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_{12} x_1 x_2 + \cdots + \beta_{123} x_1 x_2 x_3 + \cdots + \beta_{1234} x_1 x_2 x_3 x_4 + \cdots + \beta_{12345} x_1 x_2 x_3 x_4 x_5 + \epsilon \qquad (2)$$

Here *y* is the response variable, the β's are unknown parameters, x_1, x_2, x_3, x_4, x_5 represent factors A, B, C, D and E, respectively, and ϵ is a random error term. The variables x_1, x_2, x_3, x_4, x_5, are coded as 1 and −1, for the high and low levels for their respective factors. The interaction between x_1 and x_2 is denoted as $x_1 x_2$, and the other interaction effects are similarly defined. It is well known that the ordinary least squares estimates (OLS) of the β's in the Equation 2, by using Equation 3, (Wu & Hamada, 2009).

$$\hat{\beta}_{ols} = (x^T x)^{-1} x^T y \qquad (3)$$

where $y = (y_1, y_2, \dots, y_n)$, contains the responses values.

The model in terms of the observations, Equation 1, may written in matrix notation as :

$$y = X\beta + \varepsilon$$

where,

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \qquad X = \begin{bmatrix} 1 & X_{11} X_{12} \dots X_{1k} \\ 1 & X_{21} X_{22} \dots X_{1k} \\ . & . \quad . \ \dots \quad . \\ . & . \quad \dots \quad . \\ 1 & X_{n1} X_{n1} \dots X_{nk} \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}, \quad \epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix},$$

In factorial design, the design matrix x written in terms of design variable in standard order such as (for a single replicate of a 2^k design)

$$X = [\,1 \quad x_{11} \quad x_{12} \cdots \quad x_{nk}]$$

and

$$X^T Y = \begin{bmatrix} 1^T Y \\ x_{11}^T Y \\ x_{12}^T Y \\ . \\ . \\ . \\ x_{nk}^T Y \end{bmatrix}$$

and then

$$\hat{\beta} = (X^T X)^{-1} X^T Y = \frac{1}{2^k} \begin{bmatrix} 1^T Y \\ x_{11}^T Y \\ x_{12}^T Y \\ . \\ . \\ . \\ x_{nk}^T Y \end{bmatrix},$$

The predicted response is then given by:

$$\hat{Y} = X\hat{\beta} \quad ,and \;\; \hat{Y}_i = X_i^T \hat{\beta}.$$

The predicted response for any specific experiment (Montgomery, 2001) is scalar notation, the fitted model is :

$$\hat{y}_i = \hat{\beta}_0 + \sum_{j=1}^{k} \beta_j x_{ij} \text{ , i=1,2,….,n} \qquad (4)$$

4. STUDY SAMPLE

The study sample was taken from the Diwaniyah Dairy Factory in Iraq. This factory produces several types of cheese and cream and ice-cream and several types of dairy. It contains five stores used for storage of dairy products until they are marketed. The purpose of this experiment is to determine the most important factors affecting the damage of the product during the storage period, which represents the response variable. The type of the studied products is (yoghurt). Five factors were studied, suspected to have an effect on product damage during the storage period, there were the temperature, packing tray size, humidity, light intensity, and storage period. The factorial experiment design type (2^5), was chosen for identification of the main effects and interactions between these factors affecting the product during storage, with two levels for each factor, where the levels are encoded according to the rule, high level (+1), low level (-1). This study contains 5 main factors, 10 two-factor interactions, 10 three-factor interactions, 5 four-factor interactions and 1 five-factor interaction, which are described below in Table 1.

Table 1. Factors and level used in 2^5 factorial experiment

Factors	Factor Levels	
	Low level (-1)	**High level (1)**
$x_1 = A$: Temperature	<18°C	>18°C
$x_2 = B$: Humidity	<5%	>5%
$x_3 = C$: Light Intensity	Low	High
$x_4 = D$: Package size	200 ml	400 ml
$x_5 = E$: Storage period	<3days (less than 3 days)	>3 days (more than 3 days)

Source: produced by the authors

A=temperature - has two levels (low level <18 °C and high level >18 °C)
B=humidity- has two levels (low level < 5% and high level >5%)
C=light intensity- has two levels (low level and high level)
D=package size - has two levels (low level 200 ml and high level 400 ml)
E=storage period - has two levels (low level <3 days -less than 3 days- and high level >3 days -more than 3 days-)
AB=two–factor interactions between temperature and humidity
AC=two–factor interactions between temperature and light intensity
AD=two–factor interactions between temperature and package size

AE=two–factor interactions between temperature and storage period
BC=two–factor interactions between humidity and light intensity
BD=two–factor interactions between humidity and package size
BE=two–factor interactions between humidity and storage period
CD=two–factor interactions between light intensity and package size
CE=two–factor interactions between light intensity and storage period
DE=two–factor interactions between package size and storage period
ABC=three–factor interactions between temperature, humidity and light intensity
ACD=three–factor interactions between temperature, light intensity and package size
ABD=three–factor interactions between temperature, humidity and package size
BCD=three–factor interactions between humidity, light intensity and package size
ACD=three–factor interactions between temperature, light intensity and package size
ABE=three–factor interactions between temperature, humidity and storage period
ADE=three–factor interactions between temperature, package size and storage period
BCE=three–factor interactions between humidity, light intensity and storage period
BDE=three–factor interactions between humidity, package size and storage period
CDE=three–factor interactions between light intensity, package size and storage period
ACE=three–factor interactions temperature, light intensity and storage period
ABCD=four–factor interactions between temperature, humidity, light intensity and package size
ABCE=four–factor interactions between the temperature, humidity, light intensity, and storage period
ACDE=four–factor interactions between temperature, light intensity, package size and storage period
ABDE=four–factor interactions between temperature, humidity, package size and storage period
BCDE=four–factor interactions between humidity, light intensity, package size and storage period
ABCDE=five-factors interaction between temperature, humidity, light intensity, package size and storage period
We used progrom Minitab 18 for analysis of this data.

5. ANALYSIS AND DISCUSSION OF RESULTS

An important condition for factorial experiment is analysis, if the data (responses) follow the normal distribution. The analysis is also done for the purpose of knowing whether the responses follow the normal distribution or not. In order to test this data (responses) we will use the Kolmogorov-Smirnov test according to the following hypothesis.

H_0: The data distributed as normal distribution.

H_1 : The data not distributed as normal distribution.

Table 2. Results of Kolmogorov-Smirnov test

Test	Statistical test	Degree of freedom	p-value
Kolmogorov-Smirnov	0.37684	35	0.04315

Source: produced by the author, using the programing Minitab 18

The result of Kolmogorov-Smirnov test is shown in Table 1. It can be clearly seen that the p-value (0.04315), is at the least significant level at (0.05), representing evidence that the data is distributed as normal distribution. Now we have to find the main effect and interactions in order to determine the factors affecting the response variable (the damaged inventory).
We have using (OLS) method to estimation of parameters mentioned in section (3), where the obtained results are shown in the Tables 3 and 4.

Table 3. Estimates effects and coefficient for damaged product

Term	Effect	Coff	StDevcoff	t-value	Pr(> |t|)
Constant		10.722	1.072	10.01	0.001
A	5.125	2.562	1.137	3.35	**0.028**
B	0.875	0.4375	1.137	0.38	0.72
C	0.75	0.375	1.137	0.33	0.758
D	11.375	5.688	1.137	5.00	**0.007**
E	4.5	2.25	1.137	1.98	0.119
A*B	-8.375	-4.187	1.137	-3.68	**0.021**
A*C	-7.00	-3.5	1.137	-3.08	**0.037**
A*D	2.375	1.187	1.137	1.04	0.355
A*E	11.00	5.5	1.137	4.84	**0.008**
B*C	9.5	4.75	1.137	4.18	**0.014**
B*D	0.125	0.062	1.137	0.05	0.959
B*E	-3.75	-1.875	1.137	-1.65	0.174
C*D	2.25	1.125	1.137	0.99	0.378
C*E	-8.875	-4.437	1.137	-3.9	**0.017**
D*E	2.75	1.375	1.137	1.21	0.293
A*B*C	5.25	2.625	1.137	2.31	0.082
A*B*D	-6.875	-3.438	1.137	-3.02	**0.039**
A*B*E	10.5	5.25	1.137	4.62	**0.01**
A*C*D	-7.75	-3.875	1.137	-3.41	**0.027**
A*C*E	0.625	0.313	1.137	0.27	0.797
A*D*E	0.25	0.125	1.137	0.11	0.918
B*C*D	10.25	5.125	1.137	4.51	**0.011**
B*C*E	4.625	2.313	1.137	2.03	0.112
B*D*E	-5.75	-2.875	1.137	-2.53	0.065
C*D*E	-8.125	-4.063	1.137	-3.57	**0.023**

Source: produced by the author, using the programing Minitab 18

As explained in Table 3, our design can estimate all five main effects, all 10 two-factor interactions, and 10 three-factor interactions, assuming that four-factor and higher interactions are negligible. Through the obtained results, we have reached a set of results pertaining to the study of the influence of the main factors and interactions on the product (the damaged inventories).

5.1. Main effects

(A): Factor A (temperature). The p-value (0.028), being the least of significant level at (0.05). In terms of statistical point, it has a significant effect on the damage of the product during the storage period.
(B): Factor B (humidity) is non-significant, the p-value (0.720), is greater of significant level at (0.05). It has no effect on the damage of the product during the storage period.
(C): Factor C (light intensity) is non-significant, and the p-value (0.758), is greater of significant level at (0.05). It has no effect on the damage of the product during the storage period.
(D): Factor D (package size), is significant, the p-value (0.007) has the least significant level at (0.05), having significant effect on the damage of the product during the storage period.
(E): Factor E (storage period) is non-significant, the p-value (0.119), is greater of significant level at (0.05) and it has no effect on the damage of the product during the storage period.

5.2. Two-way interactions between factors

(a): A*B Two-factor interactions between the temperature and humidity is significant, with the p-value (0.021) having the least significant level at (0.05). It has significant effect on the damage of the product during the storage period.
(b): A*C Two-factor interactions between the temperature and light intensity, is significant, with the p-value (0.037) having the least significant level at (0.05).
(c): A*D Two-factor interactions between the temperature and package size, is non-significant, with the p-value (0.355), having the greatest significant level at (0.05).
(d): A*E Two-factor interactions between the temperature and storage period, is significant, with the p-value (0.008) having the least significant level at (0.05).
(e): B*C Two-factor interactions between humidity and light intensity, is significant, with the p-value (0.014) having the least significant level at (0.05).
(f): B*D Two-factor interactions between the humidity and package size, is non-significant, with the p-value (0.959) having the greatest significant level at (0.05).
(g): B*E Two-factor interactions between the humidity and storage period, is non-significant, with the p-value (0.174) having the greatest significant level at (0.05).
(h): C*D Two-factor interactions between light intensity and package size, is non-significant, with the p-value (0.378) having the greatest significant level at (0.05).
(i): C*E Two-factor interactions between light intensity and storage period, is significant, with the p-value (0.017) having the least significant level at (0.05).
(j): D*E Two-factor interactions between package size and storage period, is non-significant, with the p-value (0.293) having the greatest significant level at (0.05).

5.3.Three-way interactions between factors

(a): A*B*C Three-factor interactions between the temperature, humidity and light intensity, is non-significant, with the p-value (0.082) having the greatest significant level at (0.05).
(b): A*B*D Three-factor interactions between the temperature, humidity and package size is significant, with the p-value (0.039) having the least significant level at (0.05).
(c): A*B*E Three-factor interactions between the temperature, humidity and storage period, is significant, with the p-value (0.018), it's lest of significant level at (0.05).
(d): A*C*D Three-factor interactions between the temperature, light intensity and package size, is significant, with the p-value (0.027) having the least significant level at (0.05).
(e): A*C*E Three-factor interactions between the temperature, light intensity and storage period, is non-significant, with the p-value (0.797) having the greatest significant level at (0.05).
(f): A*D*E Three-factor interactions between the temperature, package size and storage period, is non-significant, with the p-value (0.900) having the greatest significant level at (0.05).
(g): B*C*D Three-factor interactions between the humidity , light intensity and package size, is significant, with the p-value (0.011) having the least significant level at (0.05).
(h): B*C*E Three-factor interactions between the humidity, light intensity and storage period is non-significant, with the p-value (0.112) having the greatest significant level at (0.05).
(i): B*D*E Three-factor interactions between the humidity, package size and storage period is non-significant, with the p-value (0.065) having the greatest significant level at (0.05).
(j): C*D*E Three-factor interactions between light intensity, package size and storage period, is significant, with the p-value (0.023) having the least significant level at (0.05).

Table 4. Analysis of variance for Damaged Product

Source of variance	DF	Seq SS	Adj SS	Adj MS	F	P-value
1-Main Effects	5	1262.9	1262.88	252.575	6.11	0.052
2-Way Interactions	10	3532.0	3532.00	353.200	8.54	0.027
3-Way Interactions	10	3769.0	3769.00	376.900	9.12	0.024
4-Way Interactions	5	1279.9	1279.88	255.975	6.19	0.041
5-Way Interactions	1	946.1	946.12	946.125	22.89	0.009
Residual Error	4	165.3	165.35	41.337		
Lack-of-fit test	1	147.3	147.35	147.347	24.5	0.16
Pure Error	3	18.0	18.00	6.000		
Total	35	10955.2				

Source: produced by the author, using the programing Minitab 18

Table 4 suggests that the main effects and Interactions are significant, but came in order: three-factor interaction was the first and had a significant effect on the damaged product. It can be clearly seen that the p-value (0.024) had its least significant level at (0.05), two-factor interaction came second with the p-value (0.027) having least significant level at (0.05). Finally, the main effects is non-significant, as we see the p-value (0.052). This is because the many factors are not individually important on the damaged product during the storage period, as we see in Table 3. Four-factor interactions and five-factors interactions are a significant . It is logic to be influential on the product is damaged if these factors were together. Therefore, not mentioned in Table 3 in detail. Also, we see the lack-of-fit term is non-significant with an F value of (24.5) and the P- value of (0.16). This implies that the relationship between the response variable (Damaged Product) and effects factors are linear.

6. CONCLUSIONS

In this work, we studied some factors with their interactions which affect Dairy products during storage period (before marketing), by using factorial experiments model. In this study, the appropriate model was selected to estimate the main effects and interactions that were evident by testing the absence of a reference to the relationship between the response variable (the damaged product) And the factors influencing it. Where it was a linear relationship. Through the obtained results, the most important factors affecting the products during the storage period have been identified. They were detailed as follows:

(a) Main effects: Factors A (temperature) and D (package size) are high significant and affect products during the storage period. As for the factors B (humidity), C (light intensity) and E (storage period), they were not individually significant and have no effect on product damage during the storage period.

(b) Two-way interactions between factors: Factors temperature with humidity (A*B), temperature with light intensity(A*C), temperature with storage period (A*E), humidity with light intensity period (B*C), and light intensity with storage period (C*E) are significant and have effects on the product damage during the storage period if they existed together.

(c) Three-way interactions between factors: factors temperature, humidity and package size (A*B*D), factors temperature, humidity and storage period (A*B*E), factors temperature, light intensity and storage period (A*C*D), factors humidity, light intensity period and storage period (B*C*D) and factors humidity, light intensity and storage period (C*D*E) are significant and have effects on the product damage during the storage period if they existed together.

It has been shown through the analysis of the results showed that the use of factorial experiments models with multiple regression model techniques has the potential to study the effect of factors and their common interactions on responses (dairy product damage). Also, we conclude from this study that the interactions of factors with each other have a significant impact on dairy damage during the storage period. Based on the conclusions reached, we recommend using factorial experiments models with multiple regression model in many industrial and administrative fields, such as stores management, production management and others.

REFERENCES

Douglas, C. M. (2001). *Design and analysis of experiments*. John Wiley & Sons. New York, NY, USA.

Fisher, R. A. (1926). The arrangement of field experiments. *J. Ministry Agric*, *33*, 503-513.

Mason, R. L., Gunst, R. F. & Hess, J. L. (2003). *Statistical design and analysis of experiments: with applications to engineering and science*, Vol. 474. John Wiley & Sons.

Massey, Jr. & Frank J. (1951). The Kolmogorov-Smirnov test for goodness of fit. *Journal of the American statistical Association*, *46* (253), 68-78.

Montgomery, D. C. (2001). *Design and analysis of experiments*. John Wiley & Sons. New York, 64-65.

Montgomery, D. C. (2009). *Design and Analysis of Experiments*, 7th Edition. Wiley, New York.

Williams, J. D. (1970). A regression approach to experimental design. *The Journal of Experimental Education*, *39* (2), 83-90.

Wu, C. F. J. & Hamada, M. (2009). *Experiments: planning, analysis and parameter design optimization*, 2nd edition. Wiley, New York.

Wu, C. J. & Hamada, M. (2000). *Experiments: planning. analysis, and parameter design optimization*. Wiley.

Yates, F. (1937). The design and Analysis of Factorial experiments. *Technical Commonwealth*, 35 (1-95). Bureau of soils, Harpenden.