International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452
Maths 2023; 8(1): 46-49

https://www.mathsjournal.com
Received: 23-10-2022
Accepted: 26-12-2022

Vesna Antoska Knights
Faculty or Technology and Technical Science, Veles, University St Kliment Ohridski, Bitola, Republic of North Macedonia

Jetmira Millaku
Faculty of Psychology, AAB College, Pristina, Kosovo, Republic of North Macedonia

Corresponding Author:
Vesna Antoska Knights
Faculty or Technology and Technical Science-Veles, University "St Kliment Ohridski, Bitola, Republic of North Macedonia

Three-factor experimental design as a tool in applied statistics

Vesna Antoska Knights and Jetmira Millaku

DOI: https://doi.org/10.22271/maths.2023.v8.i1a.929

Abstract
The purpose of this research is the implementation of Three-factor experimental design a as a tool in Applied Statistics in purpose of definition of indicators of the effectiveness of reducing stress levels and anxiety in students during meditation training Three-factor experimental design and yoga exercise training. Using this three-factor model three parameters were monitored: blood pressure, duration of sleeping, and smoking. This system of indicators is called a three-factor experimental design with two levels of variation (2^3) for the process of stress and anxiety. A way to obtain simplified models is using full factorial designs, which are important means to evaluate the influence of the factors on response. All 30 students-volunteers have been randomly selected, in the three-factor experiment, and analyzed, and in each case, regardless of gender, the effect of the factors. The result obtained from these analyses using the three-factor experiment were very similar for all the respondents. By applying this model, it shows highest influence has a coefficient of factor smoking, then hours of sleeping on response y (stress and anxiety). This model is a systematic method to determine the relationship between factors affecting a process and the output of that process. In other words, it is used to find cause-and-effect relationships. This information is needed to manage process inputs to optimize the output.

Keywords: Applied statistics, applied statistics as a tool in science, mathematical model, stress levels and anxiety, three-factor experimental design

1. Introduction

Statistical design of experiments refers to the process of planning the experiment so that appropriate data will be collected and analyzed by statistical methods, resulting in valid and objective conclusions. The statistical approach to experimental design is necessary if we wish to draw meaningful conclusions from the data. When the problem involves data that are subject to experimental errors, statistical methods are the only objective approach to analysis. Thus, there are two aspects to any experimental problem: the design of the experiment and the statistical analysis of the data. These two subjects are closely related because the method of analysis depends directly on the design employed. The three basic principles of experimental design are randomization, replication and blocking. We add the factorial principle to these three (Montgomery 2013) [8].

A full factorial designed experiment consists of all possible combinations of levels for all factors. The total number of experiments for studying k factors at 2-levels is 2k. A way to obtain simplified models is using full factorial designs, which are important means to evaluate the influence of the factors on the response (Antony 2014) [2]. Design of experiments has lately been applied to various research areas.

The purpose of this research is the optimal definition of indicators of the effectiveness of meditation training and yoga exercise training in reducing stress levels and anxiety in students. İt is considered a three-factor model with two levels of variation (2^3). The model is a systematic method to determine the relationship between factors affecting a process and the output of that process. In other words, it is used to find cause-and-effect relationships. This information is needed to manage process inputs to optimize the output.

Three-factor model in many fields of scientific research is used to determine the factors that cause the response for example as a indicators of obesity and weight regulation (Markovik & Knights 2022) [10].

2. Materials and Methods

2.1 Materials

This system of response for stress and anxiety will be performed on 30 voluntary students from a meditation group & yoga exercise group from Kosovo aged 18 to 30 old. Half of them are males and the other 15 are females. All 30 students attended an intensive course of yoga exercises.

A student-volunteers who's decides to be part of this research had a duty to have a diary of their sleeping hours and measure their blood pressure in the morning and in the evening. Also, in the diary, they were taking notice of how many cigarettes they were smoking per day.

2.2 Methods

Measurements for stress and anxiety are taken for conducted Beck Anxiety Inventory (BAI) Questionnaire to the students every week for a period of 8 weeks. BAI is a short list describing 21 anxiety symptoms (Beck *et al*., 1988) [3]. Internal consistency for the BAI = (Cronbach's α=0.92) Test-retest reliability (1 week) for the BAI = 0.75. (Creswell, 2008) [5]. The total score has a minimum of 0 and a maximum of 63. A score of 0 – 21 = low anxiety Score of 22 – 35 = moderate anxiety Score of 36 and above = potentially concerning levels of anxiety (Vogt, 2005) [12].

In this three-factor model the three factors are considered important to the response: blood pressure (factor X1), duration of sleeping (factor X2), and smoking (factor X3). A way to obtain simplified models is using full factorial designs, which are important means to evaluate the influence of the factors on response. We want to ascertain the relative importance of each of these factors on the response which is stress and anxiety (Y). This type of analysis will be performed in Excel in addition to using regression equations plus Cochran's test, Student t-test, and Fisher test with SPSS. In this papers are used standard statistical formulas and criteria (Beyer 2002; Sheskin 2020, Levine & Stephan 2022) [4, 11, 7].

3. Results and Discussion

The process of the experiment will, be presented in details for one student whose data is taken at random from the 30 monitored students. For all other will be presented just finally equations of model.

Table 1 is given the experimental planning matrix and results of the factorial experiment. The data for the student No1 is given for a period of 8 weeks (number of series). Number of series is determined as an 8 because of nature of experimental design, for two levels of variation: $N=n^k=2^3=8$. He has a beginning stress level of 30, using the results from the BAI Questionnaire. After an eight-week training meditation&yoga exercise, had a reduction of stress level and was achieved a total of 18 units.

Table 1: Matrix plan of the experiment for student No 1

N^0	X_1	X_2	X_3	x_1 Blood pressure (mmHg)	x_2 Duration of sleeping (h/day)	x_3 Smoking (cigarettes/day)	y_{i1}	y_{i2}	y_{av}	S_j^2
1	1	1	1	120	8	5	30	28	29	2.00
2	-1	1	1	70	8	5	28	27	27.5	0.50
3	1	-1	1	120	6	5	27	26	26.5	0.50
4	-1	-1	1	70	6	5	26	24	25	2.00
5	1	1	-1	120	8	3	24	22	23	2.00
6	-1	1	-1	70	8	3	22	21	21.5	0.50
7	1	-1	-1	120	6	3	21	19	20	2.00
8	-1	-1	-1	70	6	3	19	17	18	2.00

For each series for each patient, a combination of factors for high and low levels has been made, along with measurements of the questionnaire and we have two responses y_{i1} and y_{i2}, y_{av} is an average of those responses. S_j^2 is the variance of each response (Antony 2014) [2]. Calculated average value of variance for all variances from Table 1, is $S_j^2 = 1.43$.

To determine that the order of dispersions is considered homogeneous, it is necessary to calculate Cochran's criterion. Cochran's test is a statistical method for testing for differences between three or more matched sets of frequencies or proportions. Cochran's test is the traditional test for checking heterogeneity in meta-analyses. Using Cochran's criterion, we are testing hypotheses for reproductive experiments.

$$G_p = \frac{\max S_j^2}{\sum_{j=1}^{N} S_j^2} = 0.17 \quad (1)$$

G_p is calculated Cochran for the experiments and it is present as a ratio of maximal value of variance which is $\max S_j^2 = 15.86$ and the sum of all variances of dispersion $\sum_{j=1}^{N} S_j^2 = 1.43$, $G_{\alpha,f,N} = 0.6798$ is the value of Cochran, which have been read from table for *95%* confidence interval and degree of freedom $f = N(k-1) = 8$, where N-is the number of experiments which is 8 and k=2 is the number of levels of variation. If criteria $G_p \le G_{\alpha,f,N}$ is satisfied then statistical heterogeneity is determined, but if $G_{p.} \le G_{\alpha,f,N}$, is presented then the order of variances of dispersions is considered as homogeneous (Levine, 2014). Condition indicates the reproducibility of the experiments, and the dispersion is homogeneous. The linear three-factor model applied is given by the Equation (2) (Antony 2014) [2]:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \beta_{23} x_2 x_3 + \beta_{123} x_1 x_2 x_3 \quad (2)$$

where y is factor of stress, for the corresponding measurements yi1 and yi1, the factors x_1, x_2, x_3, their units are given in the table 1, while β_i is represent coefficients of regression and β_{ij} coefficient of interaction between factors. Some coefficients may be negligibly small or insignificant. The determination of the significance of the regression coefficients is done with the help of the Student's test criterion. In order to determine whether they are significant or not, first of all, the variance in which they are determined should be assessed:

$$S_\beta = \sqrt{\frac{S_j^2}{N}} = \sqrt{\frac{1.43}{8}} = 0.42 \quad (3)$$

The regression coefficients are significant if the condition is:

$$|\beta| \geq S_\beta \cdot t \quad (4)$$

where t is the calculated value of Student t test for all responses y_i in the Table 1, and it is t = 2.36. It means in our case regression coefficient will be significant if $|\beta| \geq 1.0024$ because $S_\beta \cdot t = 0.42 \cdot 2.36 = 1.0024$, otherwise coefficient is taken as insignificant coefficient in the regression equation or in other words will not be taken in calculations as listed in equation (2).
In Table 2 is presented the Matrix plan of the experiment for determining the coefficient of regression.

Table 2: Matrix plan of the experiment for determining the coefficient of regression

No	x_0	x_1	x_2	x_3	x_{12}	x_{13}	x_{23}	x_{123}	y_{av}
1	1	1	1	1	1	1	1	1	29
2	1	-1	1	1	-1	-1	1	-1	27.5
3	1	1	-1	1	-1	1	-1	-1	26.5
4	1	-1	-1	1	1	-1	-1	1	25
5	1	1	1	-1	1	-1	-1	-1	23
6	1	-1	1	-1	-1	1	-1	1	21.5
7	1	1	-1	-1	-1	-1	1	1	20
8	1	-1	-1	-1	1	1	1	-1	18

The results of the calculated values of the coefficients are according to the given equations are given below:

$$\beta_i = \frac{1}{N}\sum_{j=1}^{N} x_{ij} y_i \text{ and } \beta_{im} = \frac{1}{N}\sum_{j=1}^{N} x_{ij} x_{im} y_i \ (i \neq m) \quad (5)$$

$\beta_0 = 23.8125$, $\beta_1 = 0.8125$, $\beta_2 = 1.4375$, $\beta_3 = 3.1875$
$\beta_{12} = -0.0625$, $\beta_{13} = -0.0625$, $\beta_{23} = -0.1875$, $\beta_{123} = 0.0625$

From the above condition as listed in equation (4), the coefficients that do not satisfy the condition are neglected and the model as listed in equation (1) and new equation with coded variables will be:

$$y_i = 23.8125 + 1.4375 x_2 + 3.1875 x_3 \quad (6)$$

In the Equation (6), denotes the response values calculated according to the regression equation as listed in equation (2). In the Table 2 in column y_i are the response values obtained from the model by substituting the corresponding values of the variables, +1 and -1, i.e. with their values when at lower and upper levels, respectively. The values obtained from the y_{av} experiment with the values obtained from the model y are given in the Table 3.

Table 3: Calculations to determine compliance of the Fisher criterion

No.	y_{av}	y_i	y_{av}-y_i	$(y_{av} - y_i)^2$
1.	29	28.4375	0.56	0.316
2.	27.5	28.4375	-0.94	0.879
3.	26.5	25.5625	0.94	0.879
4.	25	25.5625	-0.56	0.316
5.	23	22.0625	0.94	0.879
6.	21.5	22.0625	-0.56	0.316
7.	20	19.1875	0.81	0.660
8.	18	19.1875	-1.19	1.410

The adequacy of the obtained equation as listed in equation (6), has been checked by using the Fisher criterion. The equation is adequate if the condition is meet:

$$F_p \leq F_t \quad F_p = \frac{S_{ad}^2}{S_j^2} \quad S_{ad}^2 = \frac{\sum_{j=1}^{N}(y_{av} - y_i)^2}{N-k-1} \quad (7)$$

The sum of all values in the last column is determined: $\sum_{j=1}^{N}(y_{av} - y_i)^2 = 5.56$ for the number of degrees of freedom $f = N-k-1 = 5$, so, we can determine variance of adequacy in equation: $S_{ad}^2 = \frac{5.56}{5} = 1.13$, where from table above, already is calculated an average of all variances is $S_j^2 = 1.43$, then is calculated Fisher value as listed in

equation (7): $F_p = \frac{1.13}{1.43} = 0.78$.

The value $F_t = 3.69$ is read from the table for the Fisher criteria, for confidence interval $p = 0.95$ and degree of freedom $f = 5$. The criteria $F_p \leq F_t$ is satisfied, it is concluded that model is adequate.
The next step is to determine the model in natural units, to get the response in real units it is necessary to convert the coded variables into natural variables. For that purpose, into the regression equation, we involve the Equation (8):

$$x_i = \frac{X_i - X_{i0}}{\Delta X_i} \quad (8)$$

Where: X-natural variable (factor), X_{i0}-mean level of natural variable, ΔX-the interval of change of x (standard deviation), x-coded variable (Kiemele *et al.*, 1997) [6]. The calculate values for this case are given in Table 3.

Table 4: Illustrates the two process parameters and their chosen levels for the experiment

Factor	Low level	High level	ΔX interval of change	Average
x_1 (blood pressure)	70	120	25	95
x_2 (hours of sleeping)	6	8	1	7
x_3 (smoking)	3	5	1	4

For each of the three factors, an appropriate replacement of the coded factors (small x) to natural (capital X) was performed as listed in equation (8) to the following relations:

$$x_1 = \frac{X_i - 95}{25},\ x_2 = \frac{X_i - 7}{1},\ x_3 = \frac{X_i - 4}{1} \tag{9}$$

In the equation with coded factors as listed in equation (6), a substitution is made as listed in equation (8), an equation in natural units is obtained:

$$y_i = 23.8125 + 1.4375\frac{X_2 - 7}{1} + 3.1875\frac{X_3 - 4}{1} \tag{10}$$

The equation in natural units has the following form:

$$Y_i = 1.00 + 1.4375X_2 + 3.1875X_3 \tag{11}$$

As we can see from equation (11), in this case the highest influence has coefficient of factor X3 (smoking), then is X2 (hours of sleeping) on response y, which is stress and anxiety.
The process have been repeated for all 30 volunteer students. The result obtained from these analyses using the three-factor experiment was the similar, almost the same for all of respondents, highest influence has coefficient of factor X3 (smoking), then is X2 (hours of sleeping-lack of sleeping), X1 blood pressure factor almost had no effect on the response, because blood pressure is consequence of stress and anxiety.
We compared our mathematical results with other published papers in the field of psychology and medicine for example "How cigarette smoking may increase the risk of anxiety symptoms and anxiety disorders: a critical review of biological pathways" (Moylan *et al.*, 1991) [9], evidence suggests smoking appears to increase the risk of developing increased anxiety. Numerous shared vulnerability factors have been identified that may increase the likelihood of both smoking and increased anxiety (Moylan *et al.*, 1991) [9]. Also the relationship between sleep quality, stress and academic performance among medical students (Alotaibi *et al.*, 2020) [1]. Our model gave excellent results.

4. Conclusion

The purpose of this research is the implementation of a Three-factor experimental design as a tool in Applied Statistics. In this paper, the model is used to the behavior of the response (Stress and anxiety) as a function of the factors (smoking, hours of sleeping-lack of sleeping ad blood pressure). As is presented in the results the greatest effect was smoking, the second influence is the factor of the duration of sleep. The blood pressure factor almost had no effect on the response.
This model helps, to determine whether a factor or a collection of factors, has an effect on the response, in that way is a very good tool because we can optimize the response. If students want to reduce their level of stress and anxiety they have to reduce smoking or give up on smoking and have quality sleep.

Three-factor model in many fields of scientific research is used to determine the factors that cause the response but the contribution of this work is that this model first time is implemented in psychology as determined the factors that have an effect on the response to stress and anxiety. Comparing with the published papers from medicine and psychology about the factors that influence to stress and anxiety we can confirm that our Three-factor experimental design gave excellent results.

5. References

1. Alotaibi AD, Alosaimi FM, Alajlan AA, Bin Abdulrahman KA. The relationship between sleep quality, stress and academic performance among medical students. Journal of Family & Community Medicine. 2020;27(1):23-28.
2. Antony J. Design of Experiments for Engineers and Scientists. Elsevier Ltd.; c2014. p. 63-85.
3. Beck AT, Epstein N, Brown G, Steer RA. An inventory for measuring clinical anxiety: psychometric properties. Journal of Consulting and Clinical Psychology. 1988;56(6):893-897.
4. Beyer WH. Standard Mathematical Tables. In: Zwillinger D. (Ed.). Probability and Statistics, CRC Press, Boca Raton; c2002. p. 622-734.
5. Creswell JW. Educational Research: Planning, Conducting and Evaluating, Quantitative and Qualitative Research. Pearson Education, Inc., London, United Kingdom; c2008.
6. Kiemele MJ, Schmidt SR, Berdine RJ. Basic Statistics: Tools for Continuous Improvement. Fourth ed. Colorado Springs, Air Academy Press; c1997.
7. Levine MD, Stephan FD. Even You Can Learn Statistics and Analytics: An Easy to Understand Guide to Statistics and Analytics 4rd Edition. Pearson FT Press; c2022. p. 211-248.
8. Montgomery CD. Design and Analysis of Experiments, John Wiley and Sons, New York; c2013.
9. Moylan S, Jacka FN, Julie A, Pasco JA, Berk M. How cigarette smoking may increase the risk of anxiety symptoms and anxiety disorders: A critical review of biological pathways. Brain Behavior. 2013;3(3):302-326.
10. Markovik G, Knights V. Model of optimisation of the sustainable diet indicators. Journal of Hygienic Engineering and Design. 2022;39(1):169-175.
11. Sheskin JD. Handbook of Parametric and Nonparametric Statistical Procedures. Chapman and Hall/CRC Press, New York; c2020, p. 1119-1134.
12. Vogt WP. Dictionary of Statistics & Methodology: A Nontechnical Guide for the Social Sciences. SAGE Publications; c2005. p. 10-25.