

Analyzing the Electricity Consumption Using Experimental Design Technique

S.F. Ghaderi, M.A. Azadeh and Sh. Bamdad

Research Institute of Energy Management and Planning

Department of Industrial Engineering, Faculty of Engineering, University of Tehran, Iran

Abstract: Experimental design technique is a powerful tool that is used in uncertainty cases, when a great amount of data is available. In this paper, the effects of different factors on the determination of electricity consumption are analyzed. This analysis is based on experimental design technique. The implementation of the proposed technique is shown applying electricity consumption in Iran.

Key words: ANOVA analyzes, experimental design, electricity consumption

INTRODUCTION

Due to the uncertainty of electricity consumption and the great amount of data available on electric power databases, experimental design is a powerful tool for analyzing the electricity consumption data.

Experimental design is a set of experiments that are performed on the process or system, the input data is changed by the output responses and the relations between inputs and outputs. The main goal of experimental design is to determine the variables which have the maximum effect on the responses. Initially, the experimental design was applied in the agronomy and chemistry [1-3]. The electronics industry used this method to develop the processes and products [4]. It can be seen that experimental design, is used as a main tool for statistical analysis of data in several areas. The application of experimental design technique for analyzing the variables which affect the spot price is also illustrated in [5].

The aim of this paper is to analyze the effect of different factors on determination of electricity consumption based on experimental design technique. First the mathematical model is proposed, then the application of this method of determination of electricity consumption is shown and finally a case study of the method is presented.

Factorial design: Many experiments involve a study of the effects of two or more factors. Generally for this type of experiment, it can be shown that, factorial design is most efficient. By factorial design each complete trial or replication of the experiment, all possible combinations levels of the factors are investigated. The effect of a factor is defined for as the change in response produced by a change in the level of the factor. This is frequently called a main effect because it refers to the primary factors of interest in the experiment [6].

The three factor analysis of the variance model is as follows: There are a levels of factor A , b levels of factor B and c levels of factor C .

$$y_{ijkl} = \mu + \tau_i + \beta_j + \gamma_k + (\tau\beta)_{ij} + (\tau\gamma)_{ik} + (\beta\gamma)_{jk} + (\tau\beta\gamma)_{ijk} + \varepsilon_{ijkl} \quad \begin{cases} i = 1, 2, \dots, a \\ j = 1, 2, \dots, b \\ k = 1, 2, \dots, c \\ l = 1, 2, \dots, n \end{cases} \quad (1)$$

where, μ is the total mean, τ_i is the effect of the i th level of the factor A , β_j is the effect of the j th level of the factor B , γ_k is the effect of the k th level of the factor C , $(\tau\beta)_{ij}$ is the effect of the interaction between τ_i and β_j , $(\tau\gamma)_{ik}$ is the effect of the interaction between τ_i and γ_k , $(\beta\gamma)_{jk}$ is the effect of the interaction between β_j and γ_k , $(\tau\beta\gamma)_{ijk}$ is the effect of the interaction between τ_i , β_j and γ_k and ε_{ijkl} is a random error term. Three factors are initially assumed fixed and the experiment is repeated n times, there will be nab observations, which are assumed to be normally distributed to media μ_{ij} and variance σ^2 . It is interesting to test the hypothesis about the equality of the effect of the first, second and third factors, the following tests of hypothesis are performed:

$$\begin{cases} H_0^I: & \tau_1 = \tau_2 = \dots = \tau_a = 0 \\ H_1^I: & \text{at least one } \tau_i \neq 0 \end{cases} \quad (2)$$

$$\begin{cases} H_0^J: & \beta_1 = \beta_2 = \dots = \beta_b = 0 \\ H_1^J: & \text{at least one } \beta_j \neq 0 \end{cases} \quad (3)$$

$$\begin{cases} H_0^K: & \gamma_1 = \gamma_2 = \dots = \gamma_c = 0 \\ H_1^K: & \text{at least one } \gamma_k \neq 0 \end{cases} \quad (4)$$

Corresponding Author: S.F. Ghaderi, Research Institute of Energy Management and Planning, Department of Industrial Engineering, Faculty of Engineering, University of Tehran, Iran

Table 1: The analysis of variance

Source of Variation	Sum of Squares	Degree of Freedom	Mean Square	F
A	SS _A	a-1	MS _A	F _A
B	SS _B	b-1	MS _B	F _B
C	SS _C	c-1	MS _C	F _C
AB	SS _{AB}	(a-1)(b-1)	MS _{AB}	F _{AB}
AC	SS _{AC}	(a-1)(c-1)	MS _{AC}	F _{AC}
BC	SS _{BC}	(b-1)(c-1)	MS _{BC}	F _{BC}
ABC	SS _{ABC}	(a-1)(b-1)(c-1)	MS _{ABC}	F _{ABC}
Error	SS _E	abc(n-1)	MS _E	
Total	SS _T	abcn-1		

And it is interesting to determine the interaction between factors, so, the following tests of hypothesis are formulated:

$$\begin{cases} H_0''' : (\tau\beta)_{ij} = 0 & \text{For all } i, j \\ H_1''' : \text{at least one } (\tau\beta)_{ij} \neq 0 \end{cases} \quad (5)$$

$$\begin{cases} H_0'''' : (\tau\gamma)_{ik} = 0 & \text{for all } i, k \\ H_1'''' : \text{at least one } (\tau\gamma)_{ik} \neq 0 \end{cases} \quad (6)$$

$$\begin{cases} H_0''''' : (\beta\gamma)_{jk} = 0 & \text{for all } j, k \\ H_1''''' : \text{at least one } (\beta\gamma)_{jk} \neq 0 \end{cases} \quad (7)$$

$$\begin{cases} H_0'''''' : (\tau\beta\gamma)_{ijk} = 0 & \text{for all } i, j, k \\ H_1'''''' : \text{at least one } (\tau\beta\gamma)_{ijk} \neq 0 \end{cases} \quad (8)$$

Total sum of squares SS_T is described as follows:

$$SS_T = SS_A + SS_B + SS_C + SS_{AB} + SS_{AC} + SS_{BC} + SS_{ABC} + SS_E$$

where SS_A is the sum of squares for the main effect A, SS_B is the sum of squares for the main effect B, SS_C is the sum of squares for the main effect C, SS_{AB} is the sum of squares of the interaction between factors A and B, SS_{AC} is the sum of squares of the interaction between factors A and C, SS_{BC} is the sum of squares of the interaction between factors B and C, SS_{ABC} is the sum of squares of the interaction between factors A, B and C and SS_E is the sum of squares of the errors. The tests of hypothesis are based on a comparison between the independent estimates of σ^2 provided by the division of each term of SS_T by their degree of freedom, known as mean square:

$$MS_A = \frac{SS_A}{a-1} \quad MS_B = \frac{SS_B}{b-1} \quad MS_C = \frac{SS_C}{c-1}$$

$$MS_{AB} = \frac{SS_{AB}}{(a-1)(b-1)} \quad MS_{AC} = \frac{SS_{AC}}{(a-1)(c-1)} \quad MS_{BC} = \frac{SS_{BC}}{(b-1)(c-1)}$$

$$MS_{ABC} = \frac{SS_{ABC}}{(a-1)(b-1)(c-1)} \quad MS_E = \frac{SS_E}{abc(n-1)}$$

The effect of a factor is defined by the variations in the levels of factors, that is called main effect because it refers to the primary factors. Assuming fixed factors A, B and C, the expected mean squares are:

$$E(MS_A) = \sigma^2 + \frac{bcn \sum \tau_i^2}{a-1}$$

$$E(MS_B) = \sigma^2 + \frac{acn \sum \beta_j^2}{b-1}$$

$$E(MS_C) = \sigma^2 + \frac{abn \sum \gamma_k^2}{c-1}$$

$$E(MS_{AB}) = \sigma^2 + \frac{cn \sum \sum (\tau\beta)_{ij}^2}{(a-1)(b-1)}$$

$$E(MS_{AC}) = \sigma^2 + \frac{bn \sum \sum (\tau\gamma)_{ik}^2}{(a-1)(c-1)}$$

$$E(MS_{BC}) = \sigma^2 + \frac{an \sum \sum (\beta\gamma)_{jk}^2}{(b-1)(c-1)}$$

$$E(MS_{ABC}) = \sigma^2 + \frac{n \sum \sum \sum (\tau\beta\gamma)_{ijk}^2}{(a-1)(b-1)(c-1)}$$

$$E(MS_E) = \sigma^2$$

If the null hypothesis of Eqs. (2)- (8) are true, then MS_A, MS_B, MS_C, MS_{AB}, MS_{AC}, MS_{BC}, MS_{ABC} and MS_E all estimate σ^2 . However, if there is a difference between first factor effects, then MS_A will be larger than MS_E. Similarly, this is true for the MS_B, MS_C, MS_{AB}, MS_{AC}, MS_{BC} and MS_{ABC}. So, to test the significance of three main effects and their interactions, simply divide mean square of the error mean. Large values of this ratio imply that the data do not support the null hypothesis.

If assume that the error terms \mathcal{E}_{ijkl} are normally and independently distributed with constant variance σ^2 , then each of the ratios of mean squares:

$$F_A = \frac{MS_A}{MS_E} \quad F_B = \frac{MS_B}{MS_E} \quad F_C = \frac{MS_C}{MS_E} \quad F_{AB} = \frac{MS_{AB}}{MS_E}$$

$$F_{AC} = \frac{MS_{AC}}{MS_E} \quad F_{BC} = \frac{MS_{BC}}{MS_E} \quad F_{ABC} = \frac{MS_{ABC}}{MS_E}$$

are follow an F distribution with two degrees of freedom; one degree related to the numerator term and the other related to the denominator . The critical region would be the upper tail of the F distribution; is shown in Fig. 1.

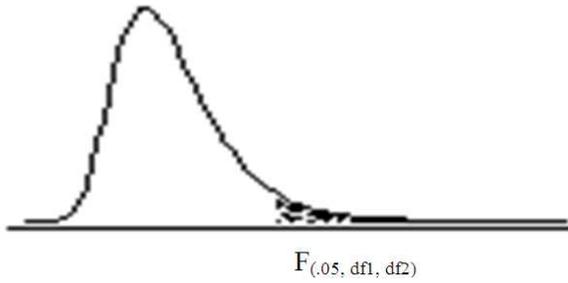


Fig. 1: F distribution

Rejecting the null hypothesis H'_0, H''_0 and H'''_0 implies that there are differences between the mean, although the exact nature of the differences is not specified. In this case, multiple comparison technique between levels is useful. Next part illustrates the application of these methods to determine the most influential factors in electricity consumption.

Electricity consumption determination: The electricity consumption data are available from January 1994, taken every hour a day, 7 days a week and 12 months a year. This information can be useful for making a response surface, which can be used to identify the value of consumption in different months of the year, different days of the week or different hours within a day.

Consider the electricity consumption is determined every hour during 24 hours for 7 days a week for 12 months a year. These data can be considered as a result of a three- factorial experiment. The first factor is the month of the year analyzed in 12 levels, the second factor is the day of the week analyzed in 7 levels and the third factor being the hour of the day which is analyzed in 24 levels, four observations per cell are selected. Each factor is previously determined, so, this is a fixed factor model. To illustrate the proposed method, the hourly electricity consumption data on Iranian from March 2003 to February 2004 is taken as an example. (The Iranian year begins on the 21th (March).

The first step is to determine of the data that satisfies the hypothesis of statistical linear model Eq. (1). The error, referred as residual is considered as the difference between the observed value y_{ijk} and the estimated value \bar{y}_{ijk} :

$$\varepsilon_{ijk} = y_{ijk} - \bar{y}_{ijk}$$

Figure 2 shows the accumulative distribution residuals from Mar. 2003 to Feb. 2004, the histogram is presented in Fig. 3.

Figure 2 and 3 shows that, the residuals are normally distributed with zero mean, hence, its representation should be a line. Next ANOVA Table is drawn from the data. Table 2 shows the results for the data from Mar.2003 to Feb.2004.

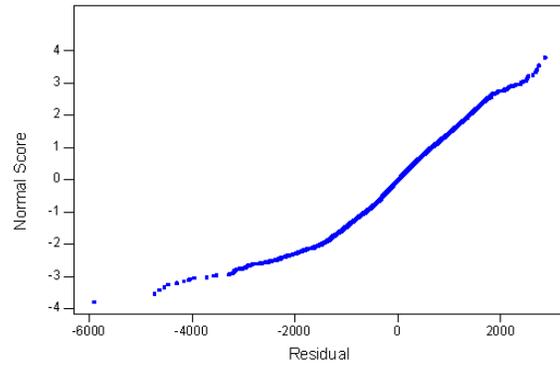


Fig. 2: Accumulative distributed residuals Mar. 2003 to Feb. 2004

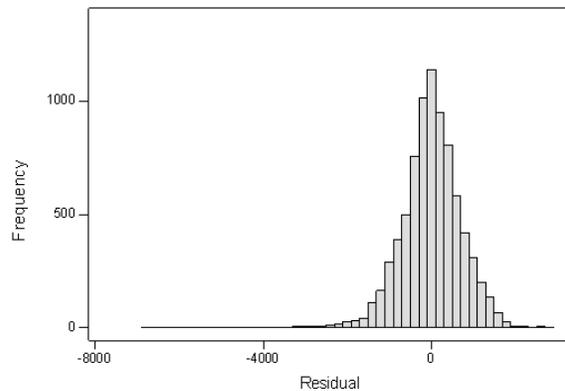


Fig. 3: Histogram Mar. 2003 to Feb 2004

From Table 2 the following results are made:

- Reject H'_0 and conclude that the month of the year affects the consumption.
- Reject H''_0 and conclude those days of the week influencing the consumption.
- Reject H'''_0 and conclude the hour of the day affects the consumption.
- Reject H''''_0 and conclude that there is an interaction between the months of the year and the day of the week.
- Reject H''''_0 and conclude that there is an interaction between the month of the year and the hour of the day.
- Reject H''''_0 and conclude that there is an interaction between the day of the week and the hour of the day.

Do not reject H''''''_0 and conclude that there is no interaction between the month of the year, the day of the week and the hour of the day.

In addition, the following values of P are obtained:

$$\begin{matrix}
 P_{months=0} & P_{days=0} & P_{hours=0} & P_{months\&days=0} \\
 P_{months\&hours=0} & P_{days\&hours=0} & P_{months\&days\&hours=1}
 \end{matrix}$$

These probabilities quantify the significance of the level of decisions. The probability that the first, second, third, fourth, fifth and sixth conclusions are wrong is near zero, whereas the probability of the seventh conclusion to be correct is almost one.

The next step is to determine if there are, months of the year, days of the week or hours of the day for which the consumption is considerably different. This is obtained using multiple comparisons based on the test proposed by Tukey and Kramer, with a trust degree of 95%. The following figures show the results. Figure 4 shows the consumption calculated for each month of the year. (Appendix A, Table A.1)

From this Fig. 4 it can be observed that the mean electricity consumption on first month is lower than the mean consumption on the other months.

Figure 5 shows the result of different days of the week. (Appendix A, Table A.2) It can be observed that mean consumption on the 7th day is considerably lower than the other days. It is mentioned that in Iran this day of the week (Friday) is a holiday.

Figure 6 shows the result for different hours of a day. (Appendix A, Table A.3) It is observed that the mean consumption is significantly greater from 7 to 11 p.m. where the residential load increases.

Now, it is possible to group months, days and hours. Month 3, 4, 5, 6 and 7 have the most consumption. Month 2, 8, 9, 10, 11, 12 have equal consumption so they are considered as normal month. In addition, the first month has the lowest consumption that is an atypical month. Days 1, 2, 3, 4, 5 and 6 are normal days and grouped as working days and the 7th day is considered as a non working day.

Hours are classified as: peak from 6 to 12 p.m., Valley form 1 to 10 a.m. and rest form 11 a.m. to 5 p.m.

Development: Let us consider as a developer, the fourth factor is the season of the year. Now add the season of the year to the model as a factor. Therefore, suppose that the electricity consumption is determined every hour during 24 hours for 7 days a week, 3 months of the season and 4 seasons of the year. These consumptions can be considered as the result of a four factorial experiment where the first factor is the season of the year analyzed at 4 levels, the second factor is the month of the season analyzed at 3 levels, the third factor is the day of the week analyzed at 7 levels and the fourth factor is the hour of the day analyzed at 24 levels. As the previous experiment, four observations per cell are selected. So the model is described as the following:

$$\begin{aligned}
 y_{ijklm} = & \mu + \tau_i + \beta_j + \gamma_k + \lambda_l + (\tau\beta)_{ij} + (\tau\gamma)_{ik} \\
 & + (\tau\lambda)_{il} + (\beta\gamma)_{jk} + (\beta\lambda)_{jl} + (\gamma\lambda)_{kl} + (\tau\beta\gamma)_{ijk} + \\
 & (\tau\beta\lambda)_{ijl} + (\tau\gamma\lambda)_{ikl} + (\beta\gamma\lambda)_{jkl} + (\tau\beta\gamma\lambda)_{ijkl} + \varepsilon_{ijklm}
 \end{aligned}$$

Next ANOVA Table is drawn from the data set. Table 3 shows the results. From Table 3, it can be seen that the season of the year influences the electricity consumption. Now it should be determined if there are seasons of the year for which the electricity consumption is considerably different. By using the test method of Tukey and Kramer, with a trust degree of 95%, we obtain Fig. 7 that shows the electricity consumption calculated for each season of the year. (Appendix A, Table A.4)

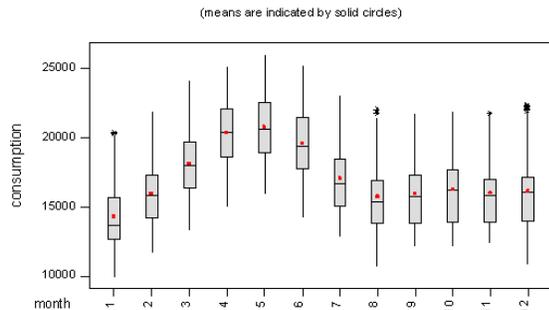


Fig. 4: Month electricity consumption comparison

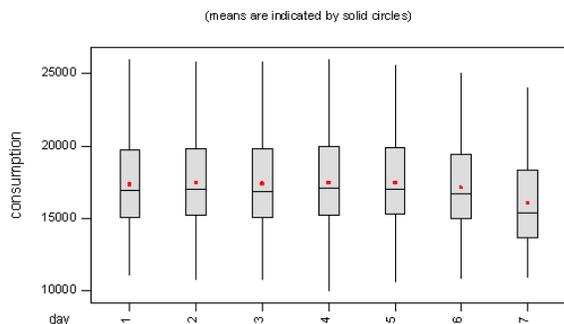


Fig. 5: Day electricity consumption comparison

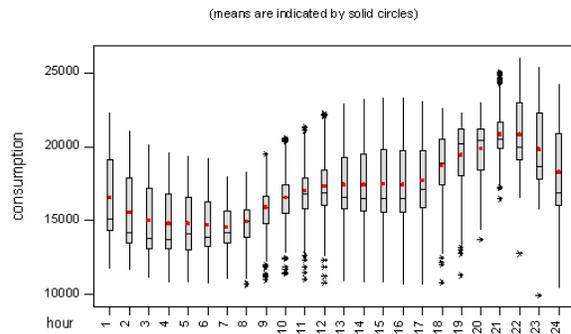


Fig. 6: Hourly electricity consumption comparison

Table 2: ANOVA, Mar. 2003 to Feb. 2004

Source of Variation	Sum of squares	Degree of freedom	Mean Square	F
Months	3. 1010E+10	11	2819123316	4015.33
Days	1890105807	6	315017635	448.69
Hours	2. 9997E+10	23	1304219351	1857.62
Month and day	415263194	66	6291867	8. 96
Month and hour	7693821428	253	30410361	43. 31
Day and hour	857728356	138	6215423	8. 85
Month and day and hour	171064218	1518	112691	0. 16
Error	4246239847	6048	702090	
Total	7. 6282E+10	8063		

Table 3: ANOVA Table

Source of Variation	Sum of Squares	Degree of Freedom	Mean Square	F
Season	2.4835E+10	3	8278177657	1.2E+04
Month	312435506	2	156217753	222.50
Day	1890105807	6	315017635	448.69
Hour	2. 9997E+10	23	1304219351	1857.62
Season and month	5863387994	6	977231332	1391.89
Season and day	126216896	18	7012050	9. 99
Season and hour	6485580460	69	93993920	133. 88
Month and day	68450319	12	5704193	8. 12
Month and hour	62322560	46	1354838	1. 93
Day and hour	857728356	138	6215423	8. 85
Season and month and day	220595979	36	6127666	8. 73
Season and month and hour	1145918408	138	8303757	11. 83
Season and day and hour	74818175	414	1800720	0.26
Month and day and hour	27084343	276	98132	0. 14
Season and month and day and hour	69161700	828	83529	0. 12
Error	4246239847	6048	702090	
Total	7.6282E+10	8063		

Table 4: ANOVA table for days of week

Source of Variation	sum of squares	degree of freedom	mean square	F	P
Day	12682155	4	3170539	0.33	0.860
Error	5. 574E+10	5755	9684820		
Total	5. 575E+10	5759			

Table 5: ANOVA table for months of the year

Source of variation	sum of squares	degree of freedom	mean square	F	P
Day	36041801	3	12013934	1.91	0.125
Error	1. 687E+10	2684	6283991		
Total	1. 690E+10	2686			

Table 6: ANOVA Table for season of the year

Source of variation	sum of squares	degree of freedom	mean square	F	P
Season	18946058	2	9473029	1.42	0.242
Error	4035E+10	6045	6675453		
Total	4037E+10	6047			

In this Fig., it can be observed that the mean electricity consumption for the second season (summer) is remarkably higher than the mean electricity consumption calculated for the other seasons. But in other seasons the electricity consumption is nearly equal. So seasons can be grouped as, the summer is peak and the others namely winter, spring and autumn is normal.

Statistical analysis: As it is shown on the model, electricity consumption is considered as a function of three factors: 1- months of year, 2- days of the week and 3- hours of the day. With regard to the Fig. 2, all of the three factors are affected on the electricity

consumption. By using multiple comparison, it could be determined that, which levels of factors are considerably different. Let us show the statistical analysis. We want to determine if we omit the levels which are considerably different, as a result the effect of the factor is still remain or not?

Analysis of the consumption versus days of the week: Figure 5 shows the result of different days of the week. It can be easily concluded that the mean consumption in 6th and 7th days is lower than the other days.

Hence, the 6th and 7th days from levels of this factor are omitted. Then the levels of the days are reduced to 5 levels (from 1st to 5th) by using pair comparison

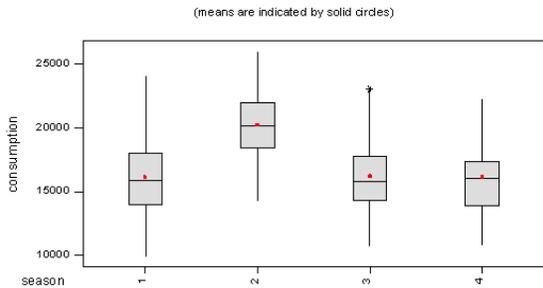


Fig. 7: Seasonally electricity consumption comparison

Between consumption and the days of the week based on the test proposed by Turkey and Kramer, analysis of the variance obtained and illustrated in Table 4.

Table 4, shows that by omitting the day 6th and the day 7th of the week, the days of the week have no effect on the consumption. This can be described that the effect of the days of the week belongs to the mentioned days of the week which the consumption is remarkably lower than other days. It can be explained that, the consumption on the day 6th and the day 7th of the week in which the consumption is remarkably lower than other days, affected on the consumption. As a result, by omitting the day 6th and 7th, we can omit the effect of factor "days" in our model.

Analysis of the consumption versus months of the year: Figure 4 shows the consumption of each month of the year. By analyzing it shows that, the effect of the month of the year is omitted only by omitting the first eight months of the year. The following Table, shows the pair comparison between consumption and the "month" of the year which the first eight months of the year are omitted and only the last four months of the year is considered.

As it is shown in Table 5, by dominating the first eight months of the year and using only the last four months of the year, the effect of the month of the year on consumption is eliminated. However, because of the eliminating over than 66 percent (8/12) of the levels of the factor it is resulted that the effect of the month of the year on consumption is not because of the special levels of the factor, but it has full random effect.

A similar statistical analysis is used for hours of the day. It could not be found any "hour" during 24 hours a day in which, by omitting that "hour" the effect of it on consumption is dominated. So, the effect of the hours of the day is random and special levels of this factor are not causing the effect on consumption.

Analyzing the consumption versus seasons of the year: Figure 7 shows the consumption of each season of the year. In this Fig. It can be observed that the consumption for summer is significantly higher than the other seasons. Now, for the statistical analysis, we are

dominating summer from the levels of the seasons. So, the season of the year is analyzed in three levels: spring, autumn and winter. By using pair comparison between consumption and seasons of the year, Table 6 is resulted.

It is obvious that the seasonal affecting on the consumption is because of high consumption in the summer.

It can be resulted from statistical analysis that the two factors, hours of the day and month of the year, have a random effect on the consumption and special levels have no role on the effect of the mentioned factors on consumption. By eliminating the summer from the levels of the seasons and 6th and 7th days of levels of the day, the effect of these two factors on consumption is omitted, it shows, that the effect of these two factors on consumption belongs to levels of these factors that the consumption is remarkably different from other levels. So, special levels of these factors influence on their effect on consumption.

CONCLUSION

The application of the experimental design to analyze the effect of different factors on electricity consumption, is illustrated in this paper. An example of this method is implemented for Iranian electricity consumption. By analyzing the result, factors can be grouped for economic purposes. Months 3,4,5,6 and 7 have the most consumption. Other months have lower consumption and considered as normal months. Days 1,2,3,4,5 and 6 are working days and the day 7th of the week is considered as a non working day. Hours are classified as: peak from 6 to 12 p.m. Valley from 1 to 10 a.m. and the rest from 11 a.m. to 5 p.m.

It can be resulted from statistical analysis that the two factors, hours of the day and month of the year, have a random effect on consumption. But special levels of two factors: days of the week and season of the year, that the consumption is remarkably different from other levels, influence on their effect on consumption.

APPENDIX A

Table A.1: Consumption versus Month comparison

Level	N	Mean	StDev	-----+-----+-----
1	672	14292	2270	*)
2	672	15980	2199	(*)
3	672	18148	2212	(*)
4	672	20340	2234	(*)
5	672	20757	2239	(*)
6	672	19582	2413	(*)
7	672	17052	2434	(*)
8	672	15751	2389	(*)
9	672	15959	2524	(*)
10	672	16257	2540	(*)
11	672	16006	2436	(*)
12	672	16129	2526	(*)
				-----+-----+-----
				16000 18000 20000

