Use of Design of Experiment in Media Industry

Anushree Saha[1], Archana Jaikumar Nair[1], Sanika Mhamunkar[1], Sneha M[1]

[1]M.Sc. Applied Statistics, Symbiosis Statistical Institute, Pune

*Corresponding author Email ID: anushree.saha@ssi.edu.in , archana.nair@ssi.edu.in , sanika.mhamunkar@ssi.edu.in , sneha.m@ssi.edu.in

ABSTRACT:

Media are the communication outlets or tools used to store and deliver information or data. The term refers to components of the mass media communications industry, such as print media, publishing, the news media, photography, cinema, broadcasting (radio and television), digital media, and advertising. The influence of media on individuals and society is increasing. The paper observes how DOE, a powerful data collection and analysis tool, can be used in variety media studies research. It considers the types of research questions experiments can address, how the experiment can demonstrate causal relationships between variables, common experimental designs, and interpretation.

The paper focuses on solving 3 case studies using Completely Randomised Design, Random Block Design and Taguchi method respectively, to analyse 3 different media related topics. The study uses MS Excel, R programming, Minitab to perform the analysis.

INTRODUCTION:

Media is defined as a tool or communication outlet that is used to store and deliver information to people. The media industry is a business model based on communicating information, art and entertainment to a large audience. The media industry encompasses all businesses that allow information to be shared. The way that this information is shared varies massively, with some examples of media including radio, television, newspapers, social media, video games, film and music. Media can be split into two categories: traditional and digital media.
Traditional media is any form of mass communication that occurred before the introduction of the internet, e.g. radio, newspapers, television and billboards. Traditional media is often referred to as old media.
Digital media is any form of media introduced after the internet that uses an electronic device for distribution. It is quite a broad term, and encompasses media types such as social media, podcasts, online games, blogs and YouTube. Digital media is often referred to as new media.

Design of Experiment:
Design of experiments (DOE) is defined as a branch of applied statistics that deals with planning, conducting, analysing, and interpreting controlled tests to evaluate the factors that control the value of a parameter or group of parameters. DOE is a powerful data collection and analysis tool that can be used in a variety of experimental situations.
It allows for multiple input factors to be manipulated, determining their effect on a desired output (response). By manipulating multiple inputs at the same time, DOE can identify important interactions that may be missed when experimenting with one factor at a time. All

possible combinations can be investigated (full factorial) or only a portion of the possible combinations (fractional factorial).
A strategically planned and executed experiment may provide a great deal of information about the effect on a response variable due to one or more factors. Many experiments involve holding certain factors constant and altering the levels of another variable. This "one factor at a time" (OFAT) approach to process knowledge is, however, inefficient when compared with changing factor levels simultaneously. Fisher demonstrated how taking the time to seriously consider the design and execution of an experiment before trying it helped avoid frequently encountered problems in analysis.

Completely randomized design(CRD):

A completely randomized design is probably the simplest experimental design, in terms of data analysis and convenience. With this design, subjects are randomly assigned to treatments.
A completely randomized design relies on randomization to control for the effects of extraneous variables. The experimenter assumes that, on average, extraneous factors will affect treatment conditions equally; so any significant differences between conditions can fairly be attributed to the independent variable.

Randomized Block Design(RBD):

A randomized block design is an experimental design where the experimental units are in groups called blocks. The treatments are randomly allocated to the experimental units inside each block. When all treatments appear at least once in each block, we have a completely randomized block design. Otherwise, we have an incomplete randomized block design.
This kind of design is used to minimize the effects of systematic error. If the experimenter focuses exclusively on the differences between treatments, the effects due to variations between the different blocks should be eliminated.

Taguchi Method:

The Taguchi method involves reducing the variation in a process through robust design of experiments. The overall objective of the method is to produce high quality product at low cost to the manufacturer. The Taguchi method was developed by Dr. Genichi Taguchi of Japan who maintained that variation. Taguchi developed a method for designing experiments to investigate how different parameters affect the mean and variance of a process performance characteristic that defines how well the process is functioning. The experimental design proposed by Taguchi involves using orthogonal arrays to organize the parameters affecting the process and the levels at which they should be varied. Instead of having to test all possible combinations like the factorial design, the Taguchi method tests pairs of combinations. This allows for the collection of the necessary data to determine which factors most affect product quality with a minimum amount of experimentation, thus saving time and resources.

The Taguchi method is best used when there is an intermediate number of variables (3 to 50), few interactions between variables, and when only a few variables contribute significantly.
The Taguchi arrays can be derived or looked up. Small arrays can be drawn out manually; large arrays can be derived from deterministic algorithms. Analysis of variance on the collected data from the Taguchi design of experiments can be used to select new parameter values to optimize the performance characteristic. The data from the arrays can be analysed by plotting the data and performing a visual analysis, ANOVA, bin yield and Fisher's exact test, or Chi-squared test to test significance.

TERMINOLOGIES:

Key concepts in creating a designed experiment include blocking, randomization, and replication.

- **Blocking:** When randomizing a factor is impossible or too costly, blocking lets you restrict randomization by carrying out all of the trials with one setting of the factor and then all the trials with the other setting.
- **Randomization:** Refers to the order in which the trials of an experiment are performed. A randomized sequence helps eliminate effects of unknown or uncontrolled variables.
- **Replication:** Repetition of a complete experimental treatment, including the setup.
- **Hypothesis testing** is an act in statistics whereby an analyst tests an assumption regarding a population parameter. The methodology employed by the analyst depends on the nature of the data used and the reason for the analysis. Hypothesis testing is used to assess the plausibility of a hypothesis by using sample data.
- **Factors:** The design and analysis of experiments revolves around the understanding of the effects of different variables on other variable(s). In mathematical jargon, the objective is to establish a *cause-and-effect* relationship between a number of *independent variables* and a *dependent variable* of interest. The dependent variable, in the context of DOE, is called the *response*, and the independent variables are called *factors*.
- **Levels:** Experiments are run at different factor values, called *levels*.

OBJECTIVES OF THE STUDY:

1. To analyse the effect of language on TV shows TRP ratings
2. To analyse how the different types of media and/or different industries would affect the expense in advertisement
3. To analyse which form of media advertisement has a greater effect on sales of a product.

PROBLEM STATEMENT:

1. To study the effect different languages have on the TRP rating of the same show
2. To study the effect on advertisement expenses due to different forms of media and different industries.
3. To study the effect on sales of 4 companies considering their budget for advertisement on 3 different media platforms viz TV, Radio, and Social media.

LITERATURE REVIEW:

[1] P. Sivaiah et al has depicted the culling of optimum cutting conditions in turning of 17-4 precipitation hardened stainless steel (PH SS) using Taguchi optimization method. [2] Khushi Padwe et al studied the effect of several factors affecting consumer buying behaviour towards FMCG products and to examine the financial performance of selected FMCG companies in India using discriminatory power of financial ratios.

METHODOLOGY:

1. **Source of Data:** The three different datasets that were obtained from Kaggle for the study are as follows:

   Case Study 1 - 'Bigg Boss India series - Telugu/Tamil/Kannada/Hindi/Malayalam/Marathi/Bengali' data set was analysed using Completely Randomized Design.

   Case Study 2 - 'Advertisement Expenses by Companies' data set was analyzed using Randomized Block Design.

   Case Study 3 - 'Traditional Media and Social Media Advertising on Sales performance' data set was analyzed using Taguchi Method.

2. **Tools of Analysis for data analysis:** ANOVA, Completely Randomized Design, Randomised Block Design, Taguchi Method were used to analyze the data sets.

   Case Study 1 - Microsoft Excel software was used to obtain a one-way ANOVA table for Completely Randomized Design.

   Case Study 2 - R studio software was used to obtain a two-way ANOVA table for Randomized Block Design.

   Case Study 3 - Minitab software was used to obtain Response tables and Main Effect plots for means and S/N ratio as well as ANOVA for s/n ratio.

DATA ANALYSIS:

Case Study 1: Analyse the effect of language on TV shows TRP ratings

The dataset was about the TRP ratings of a TV show which was aired in different languages. Completely randomized design was used to analyse the data as it allows complete flexibility as any number of treatments and replicates may be used. The study consisted of 5 replicates and 3 treatments. The treatment was considered to be language (namely Hindi, Tamil, Telugu) (see dataset 1).

	Hindi	Tamil	Telugu
1	2.72	7.8	15
2	2.89	8.2	15
3	2.83	8.8	18.1
4	4.95	15.6	20.1
5	4.19	18.6	18.2

Data set 1

The Hypothesis for the test is given as

H_0: There is no significant difference in TRP due to treatment

V/S

H_1: There is a significant difference in TRP due to treatment

The Decision Rule is given as

If P-value$\leq$ 0.05 (5% significance), we reject H_0.

If P-value> 0.05 (5% significance) then, we fail to reject H_0.

One-way ANOVA was processed using the Data Analysis tool pack in MS Excel, the results of which can be seen in table 3.2. From the table 1.1 we can see that the P-value calculated is 0.000070564 which is less than 0.05. Hence, the null hypothesis is rejected.

ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	480.1712533	2	240.0856267	23.51576627	7.0564E-05	3.885293835
Within Groups	122.51472	12	10.20956			
Total	602.6859733	14				

Table (1.1) – ANOVA table

Case Study 2: Analyse how the different types of media and/or different industries would affect the expense in advertisement

The data set used is - ‘Advertisement Expenses by Companies’.

	Tv	Radio	Print
Auto	144426.5	10251.4	244346.1
Food & Beverages	13305.05	5078.15	9945.45
Personal Healthcare	5176	2784	378855
Retail	16813.75	3669.359	36253.77
Telecom/Internet Service	361496.3	13425.46	15432.15
Textiles/Clothing	26861.5	8136	23931

Dataset 2

The data sets consist of various Industrial Sectors like Automobiles, Food and Beverages, Personal Healthcare, Retail, Telecommunications/Internet Service, and Textiles/Clothing in the rows and different Forms of Media i.e., Television, Radio, and Print in the columns.

The advertisement expenditure of different industrial sectors in various forms of media are given in the data. The Problem Statement was to study the effect on advertisement expense due to different forms of media and different industries. The blocks are different sectors of industries - Automobiles, Food and Beverages, Personal Healthcare, Retail, Telecommunications/Internet Service, and Textiles/Clothing. The Treatments are the Spending on Television, Radio, and Print.

The Hypothesis for the test is given as

For Treatments:

H_0: There is no significant difference in advertisement spending due to the types of Media.

V/S

H_1: There is a significant difference in advertisement spending due to at least one type of Media.

For Blocks:

H_0: There is no significant difference in advertisement spending due to the Industrial Sectors.
V/S
H_1: There is a significant difference in advertisement spending due to at least one type of Industrial Sector.

The Decision Rule is given as

If P-value$\leq$ 0.05 (5% significance), we reject H_0.

If P-value> 0.05 (5% significance) then, we fail to reject H_0.

Output

```
> av = aov(r ~ treatments + blocks, data=df2)
> summary(av)
            Df    Sum Sq   Mean Sq F value Pr(>F)
treatments   2 4.100e+10 2.050e+10   1.270  0.322
blocks       5 5.941e+10 1.188e+10   0.736  0.613
Residuals   10 1.614e+11 1.614e+10
```

Table (2.1) – ANOVA table

The software R studio was used to obtain ANOVA table for the Randomized Block Design. As seen in table 2.1, the P-value for the Treatments as 0.322 and the P-value for the Blocks as 0.613.

Case Study 3: analyse which form of media advertisement has a greater effect on sales of a product

The dataset takes into account the advertising budgets of four distinct companies for three different media platforms: television, radio, and social media. The impact on the individual companies' sales is investigated.

TV	Radio	Social Media	Sales 1	Sales 2	Sales 3	Sales 4
0	0	0	469.8	333	308.1	467.9
1	0	0	597.1	495.4	424.8	365
0	1	0	497.3	465.7	394.2	489.3
1	1	0	636.7	561.3	672.1	479.2
0	0	1	504.1	360.1	433.6	499.8
1	0	1	429.3	737.1	429.3	507.6
0	1	1	820.8	488.1	539.4	567.5
1	1	1	807.8	658.8	771	669.5

Dataset 3

The Taguchi method was used to analyse the data from Case Study 3. It is a method of ensuring good performance in the design stage of products or processes. Taguchi analysis was done using Minitab software tool and the response table for means and S/N ratios along with their plots and analysis of variance (ANOVA) for S/N ratio results were obtained and presented in the forth coming discussions.

The study considers 3 factors namely TV, Radio, and Social Media at two levels viz High budget (1) and Low budget (0). Taguchi method was used on a 2^ (3-1) (i.e. half) fractional factorial design with an orthogonal array of eight factor-level combinations for four different firm sales.

Means for factor level combination:

Minitab calculates a separate mean for each combination of control factor levels in the design.

Response Table for Means

Level	TV	Radio	Social Media
1	477.4	460.1	478.6
2	577.6	594.9	576.5
Delta	100.2	134.8	97.9
Rank	2	1	3

Table (3.1) - Response table for Means

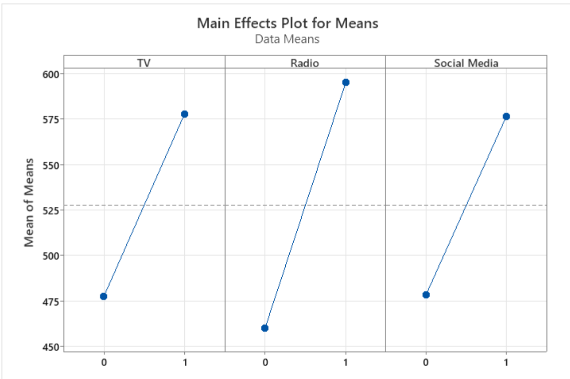

Fig (3.1) - Main effect plots for means

Fig 3.1 illustrates the main effect for means of various advertising budgets on sales for each of the factors.

Results:

Table 3.1 obtains means for each factor level combinations. For each of the three factors, the delta values represent the difference of means between levels viz 1-Low Budget and 2-High Budget. The factor with the biggest delta value has the most impact on the response variable, which in this case is sales performance. Radio has the biggest delta value - 134.8, followed by television and social media, as shown in the table. As a result, Radio is ranked the highest.

From Fig3.1 it is visible that Radio has the highest impact on sale performance. Although, because delta values do not vary greatly, we cannot draw any definitive conclusions from the major effect of means plot. Therefore, S/N ratio for each factor level combination is calculated further for more robust conclusions.

S/N ratio for factor level combination:

Genichi Taguchi used a loss function; it is a difference of experimental value and target value which is again converted into the S/N ratio. S/N ratio defined as the ratio of mean to standard deviation.

Signal to noise ratio in Minitab was set to default i.e. for nominal is the best. The signal-to-noise (S/N) ratio is calculated for each factor level combination. The formula for the nominal-is-best (II) S/N ratio using base 10 log is:

$$\frac{S}{N} = 10 * \log\left(\frac{\overline{Y}^2}{s^2}\right)$$

Response Table for Signal to Noise Ratios

Nominal is best (10×Log10(Ybar^2/s^2))

Level	TV	Radio	Social Media
1	15.43	13.57	15.82
2	15.30	17.16	14.91
Delta	0.13	3.59	0.91
Rank	3	1	2

Table (3.2) - Response table for S/N ratio

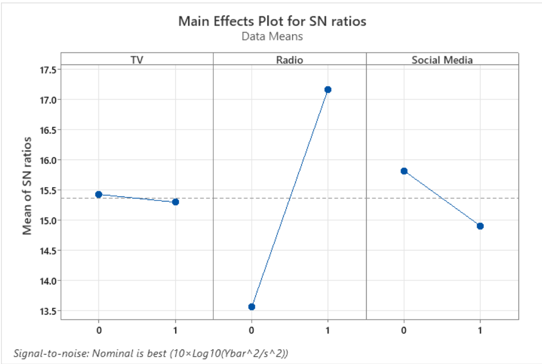

Fig (3.2) - Main effect plots for S/N ratio

Fig 3.2 illustrates the main effect for S/N ratios of various advertising budgets on sales for each of the factors.

Results:

Table 3.2 obtains S/N ratios for each factor level combinations. For each of the three factors, the delta values represent the difference of S/N ratios between levels viz 1-Low Budget and 2-High Budget. The factor with the biggest delta value has the most impact on the response variable, which in this case is sales performance. Radio has the biggest delta value - 3.59. As a result, Radio is ranked the highest. From Fig 3.2 it is visible that Radio has the highest impact on sales performance, TV has almost negligible impact and Social media has some effect on the sales performance.

ANOVA for SN ratios:

Analysis of Variance for SN ratios

Source	DF	Seq SS	Adj SS	Adj MS	F	P
TV	1	0.0332	0.0332	0.0332	0.00	0.962
Radio	1	25.7570	25.7570	25.7570	1.99	0.231
Social Media	1	1.6511	1.6511	1.6511	0.13	0.739
Residual Error	4	51.8195	51.8195	12.9549		
Total	7	79.2609				

Table (3.3)- ANOVA table for S/N ratio

ANOVA was used to check the contribution of each of the factors on the sales performance of the 4 companies. It was calculated using the sum of squares for each factors:
Seq SS/ Total SS * 100

TV = (0.0332/79.2609) *100 = 0.041%

Radio = (25.7570/79.2609) *100 = 32.49%

Social Media = (1.6511/79.2609) *100 = 2.08%

According to these calculations, Radio has contributed 32.49%, Social Media has contributed 2.08% and TV has contributed 0.041% to the sales performance. The rest is residual error.

CONCLUSION:

We used 3 experimental design on appropriate data sets and found the following conclusions

1. Case Study 1:
   Since the P-value calculated was less than 0.05 then, there is a significant difference in the mean treatment effect due to the treatment. That is there is a difference in the TRP ratings of the show due to the change in languages.

2. Case Study 2:
   The P-value for Treatments is greater than 0.05, hence we fail to reject the null hypothesis H_0 and conclude that there is no significant difference in advertisement spending due to different types of media.
   The P-value for Blocks is greater than 0.05, hence we fail to reject the null hypothesis H_0 and conclude that there is no significant difference in advertisement spending due to different Industrial sectors.
   Hence, there is no significant effect on advertisement expenses due to different forms of media and different industries.

3. Case Study 3:
   According to the delta values of means and S/N ratios of factor level combinations, Radio, ranks the highest. This implies that the change in Budgets of Radio Advertising has the most significant effect on sales. From the table 3.3, the contribution of each factor on sales is calculated, and it can be seen that Radio contributes the highest to the sales performance with 32.49%, followed by Social Media and TV. As a result, the study concludes that Radio advertisement has the highest influence on sales performance. The sales performance changes significantly as the budget for Radio advertisement changes.

LIMITATIONS & FUTURE SCOPE:

1. If the data is sufficient, the study might be further investigated by comparing different types of experimental designs.
2. Due to a lack of data, the study uses the Taguchi method with fractional factorial approach. The Taguchi approach can be investigated further for full factorial designs

REFERENCE

- Research Paper:
  [1] Sivaiah, P. and Chakradhar, D., 2019. Modeling and optimization of sustainable manufacturing process in machining of 17-4 PH stainless steel. *Measurement*, *134*, pp.142-152.
  [2] Khushi Padwe, Mahak Saxena, Mayuri Sandge, Ritu Bhasin, and Shreeman Mishra. Use of Design of Experiments in Fast Moving Consumer Goods
  [3] Design And Analysis Of Experiments - Montgomery and Douglas C

- Web sites:
  1. https://asq.org/quality-resources/design-of-experiments
  2. https://www.ndsu.edu/faculty/horsley/CRD.pdf
  3. https://academic.macewan.ca/burok/Stat252/notes/RBD.pdf
  4. http://www2.stat.duke.edu/~banks/111-lectures.dir/lect21.pdf
  5. https://www.smartcapitalmind.com/what-is-the-media-industry.htm
  6. https://support.minitab.com/en-us/minitab/18/help-and-how-to/modeling-statistics/doe/supporting-topics/taguchi-designs/taguchi-designs/
  7. https://support.minitab.com/en-us/minitab/18/help-and-how-to/modeling-statistics/doe/how-to/taguchi/analyze-taguchi-design/before-you-start/data-considerations/
  8. https://eng.libretexts.org/Bookshelves/Industrial_and_Systems_Engineering/Book%3A_Chemical_Process_Dynamics_and_Controls_(Woolf)/14%3A_Design_of_Experiments/14.01%3A_Design_of_Experiments_via_Taguchi_Methods_-_Orthogonal_Arrays

APPENDIX:

R-Codes:

```
df2 = read.table(choose.files(), header=TRUE)
df2
r = c(t(as.matrix(df2))) # response data
r
f = c("Tv", "Radio", "Print") # treatment levels
k = 3 # number of treatment levels
n = 6 # number of control blocks
treatments = gl(k, 1, n*k, factor(f))# matching treatment
treatments
blocks = gl(n, k, k*n)# blocking factor
blocks
av = aov(r ~ treatments + blocks, data=df2)
summary(av)
```