University of South Carolina
Scholar Commons

Senior Theses | Honors College

Fall 2018

Statistical Design Of Experiment Techniques in Manufacturing

Caroline M. Kerfonta
kerfontc@email.sc.edu

Follow this and additional works at: https://scholarcommons.sc.edu/senior_theses

Part of the Applied Statistics Commons, and the Design of Experiments and Sample Surveys Commons

Recommended Citation

Kerfonta, Caroline M., "Statistical Design Of Experiment Techniques in Manufacturing" (2018). *Senior Theses*. 263.
https://scholarcommons.sc.edu/senior_theses/263

This Thesis is brought to you by the Honors College at Scholar Commons. It has been accepted for inclusion in Senior Theses by an authorized administrator of Scholar Commons. For more information, please contact dillarda@mailbox.sc.edu.

STATISTICAL DESIGN OF EXPERIMENT TECHNIQUES IN MANUFACTURING

By

Caroline M. Kerfonta

Submitted in Partial Fulfillment
of the Requirements for
Graduation with Honors from the
South Carolina Honors College

December 2018

Approved:

Edsel Peña Ph.D
Director of Thesis

John Grego Ph.D
Second Reader

Steve Lynn, Dean
For South Carolina Honors College

Table of Contents

Dedication

For Dorothy and Robert Hutchings

And

For Mary and Thomas Kerfonta

Summary

This thesis is a discussion and overview of the basic techniques and uses of statistical design of experiments in manufacturing. It will introduce and discuss three different techniques. The three techniques that will be discussed are completely randomized design, randomized block design, and factorial design. Along with discussion of these three types, this thesis will explain my background, explore the history of statistical experiments in manufacturing, and discuss my future work.

The three techniques will each be explained, with their uses and constraints. This explanation will be followed by a real world example of the technique being used. Then, sample R code for the technique will be given and explained.

In addition to explanation of the three techniques, the history of statistical experiments in manufacturing and a discussion about the techniques will be included in the thesis. The work of W. Edwards Deming will be discussed because he helped to make statistical experiments in manufacturing popular. It is important to choose the correct design. Therefore, I will discuss the consequences of not choosing the correct technique.

After analysis and discussion, the thesis will conclude with an explanation of my future work and a brief synopsis of the thesis.

Abstract

There are many statistical techniques used to design experiments. These techniques are used in many different fields. This thesis will focus on the use of the three most common techniques used to design statistical experiments in manufacturing.

The three techniques that will be investigated are completely randomized design, randomized block design, and factorial design. These techniques will be compared, contrasted, and explained. Research examples will be presented along with sample R code for each technique. These examples will be accompanied by analysis of the techniques as well as an overview of the uses and history of experiments in manufacturing.

Introduction

Statistics is a science that is used in many different fields and industries. There are many techniques within the field of statistics that can be used to solve problems. One of these techniques is designing experiments so that they properly assess the questions that are being asked.

Design of experiments is used throughout many industries and research areas. This thesis will focus on design of experiment techniques and examples from automobile and aircraft manufacturing, research, and factories. Manufacturing and factories have a variety of problems and questions that can be solved or improved through experimentation. It is important to properly design these experiments in order to explore the problem. When designing an experiment, it is important to control certain factors so that the factor of interest can be investigated. In manufacturing processes, there are often many different factors and effects occurring at once. Controlling all the possible factors to limit their effects on the study is very important. There are many different techniques that can be used to design the experiments and control unrelated factors.

The three techniques that will be thoroughly explained and detailed in this paper are completely randomized design, randomized block design, and factorial design. These three techniques are the most commonly used and are the basis for the more complex design methods. All three types use randomization and share some techniques. Factorial design is an example of this, it is a more complex type of completely randomized design. Each of these types will be introduced, type of analysis used in each will be explained, and the uses in manufacturing will be detailed and critiqued. In addition, examples of their uses in manufacturing from past studies and literature will be presented and discussed.

For the examples of the analysis of each type of design, the statistical software R will be used. Code and output for each type will be provided and explained. For sample analysis and coding the R built-in dataset called "mtcars" will be used. This dataset uses data from the 1974 *Motor Trend* magazine. Its variables include fuel use (mpg), number of gears, weight, and axle ratio.

This thesis will also include a discussion about the types of design. It will compare and contrast them and their uses in manufacturing. It will also offer insight into possible improvements for future experimental designs and experiments. The conclusion will include future work and research on statistical design of experiment techniques.

My Interest/Background

My name is Caroline Kerfonta and I am a senior statistics major at the University of South Carolina. I will be graduating with my Bachelor of Science degree in December 2018. I chose to write my South Carolina Honors College Senior Thesis on the statistical design of experiment techniques used in manufacturing because of the experiences and opportunities that I had while interning with the Boeing Company in North Charleston, SC.

My position was with the 787 Finance Operations team, but I was able to interact with many different groups and individuals with diverse backgrounds. During my time at Boeing I was able to meet and interact with members of the Applied Mathematics group which is a part of Boeing Research and Technology. Through my conversions with this group, I became interested in the techniques behind designing experiments and the uses in manufacturing.

I am particularly interested in the manufacturing, processes, and research behind aircraft production because of the time that I spent working at Boeing. However, many of the examples that I have found in existing literature and have included in this thesis focus on automobiles and automobile production. The techniques and problems researched in the literature are very similar and relate to those used in aircraft production.

History of Statistics in Industry: W. Edwards Deming

Dr. William Edwards Deming was a American academic and researcher who developed techniques and processes for quality control in manufacturing. Deming used statistical methods to analyze the causes of defects in manufacturing, how to correct the defects, and to measure the outcomes of the corrections.

During his life, Deming wrote many books and papers and consulted and worked with many companies and researchers. His methods and philosophies first became popular with Japanese manufacturers. His work helped Japan to grow economically after the Second World War. Later, his work also became popular in the United States.

His method focused on fourteen points. These fourteen points create a positive environment in which production can become more efficient and more accurate. When used, Deming's points increase profits, decrease waste, and improve accuracy in industry.

Deming's points and theories became a foundation for the use of statistical analysis in industry and manufacturing. His work proved that the outcomes of experimenting with and adjusting production processes were worth the time and effort put into them. Deming's researcher showed that through improvements based upon experimentation, production can be increased and become more efficient.

The techniques and research that Deming conducted, laid a foundation for the examples and processes looked and described in this thesis.

Definition of Manufacturing

Merriam-Webster's dictionary defines manufacturing as "to make a product suitable for use" and "to produce according to an organized plan and with a division of labor". In this paper, manufacturing will refer to the processes used to produce automobiles, aircraft, and similar vehicles. This includes the research to improve the vehicle as well as the production and after production testing of the vehicle.

Examples of experiments in Manufacturing

There are many different types of experiments in manufacturing. They are incorporated during and along all of the steps of the manufacturing process. From materials testing before production begins, to production optimization, to final product testing, experiments help to improve the process and the product. These improvements can increase the safety and quality of the product and decrease the time and costs of production.

Examples of experiments in manufacturing include the testing of airplane wing flexibility, durability of car tires, improvements on production line processes and timing, and safety and ergonomics of workers. During all of these experiments, statistics can be used to analyze and determine the significance of the outcomes.

In order for the Boeing 787 airplane to be able to be produced and flown commercially, it had to pass many tests for safety and quality. One of the many tests that this airplane had to pass was the wing flex test. During this test, the wings were flexed upward by about

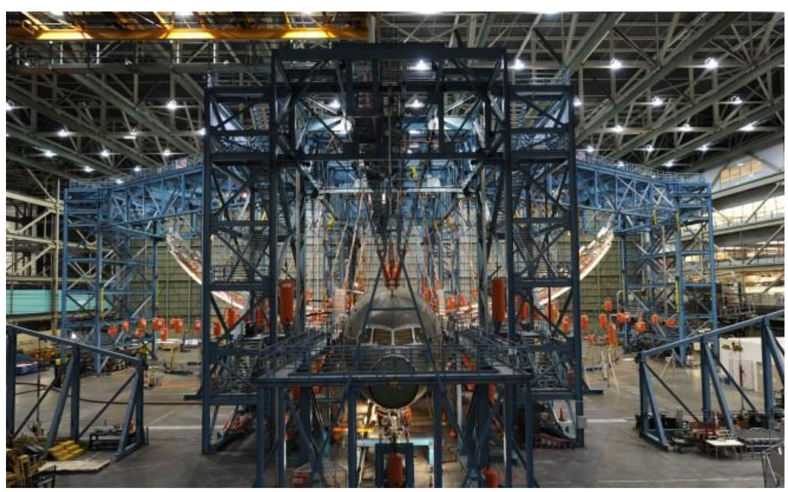

Figure 1: Boeing 787 Wing Flex Test

25 feet. This test is commonly done on aircraft. It is done in order to prove that the aircraft can carry the load it will be carrying and withstand the forces that act on it while it is flying. If the test of the wings fails, design must be reconsidered and redone before the airplane can fly. Statistics can be used during tests like these to determine the likelihood that the wing will break under different amounts of force. The wing test of the 787 is pictured to the above.

Completely Randomized Design

The most simple way to design an experiment is to use a completely randomized design. In this type of design, the treatments are randomly assigned to the subjects or objects in the study. When using this type of design, certain factors in the experiment are controlled. This is done to reduce error due to variation. However, not all factors can be controlled, in these cases, randomness is used to attempt to reduce error.

When an experiment uses completely randomized design, the objects or subjects being tested only receive one of the treatments. Typically, the number of objects in each treatment group are the same. This type of design produced easily interpreted results. It is easier to design and carry out because only one treatment is applied to each object. The relationships between the treatments are not being analyzed, only the treatment on the objects themselves. This example shows the simplest form of completely randomized design. Factorial design will be discussed later on, it is a more complex form of completely randomized design, where more than one treatment is used.

Completely Randomized Design Example

Completely randomized design was utilized by Jerkovic et al. in the 2010 study entitled, “Study of the Abrasion Resistance in the Upholstery of Automobile Seats”. This study compared and contrasted four abrasion testers that are used to test textiles in automobile manufacturing. This study was important to the manufacturing process of the seats and automobiles because seats must be safe for the passengers and driver, as well as being an important selling point for the customers.

For their completely randomized design, the researchers replicated the test five times for each of the four abrasion testers. They were using the abrasion testers to study the weight loss of the upholstery due to abrasion. The tests controlled the type of abrasive paper used, the directions of the samples, and the thickness of the upholstery sample. Randomness is used in the order that the experiments were carried out. For example, if the experiments were not randomized, all of the tests on the Martindale machine would be carried out first. However, with randomness, the order of machine usage is randomized. This controls for possible unknown differences in the upholstery samples or timing and usage of the machines overtime.

Once the experiment was conducted, an ANOVA was used to test whether or not the type of abrasion tester had an effect on the weight loss of the upholstery. The F statistics calculated from the experiment was very high, signaling that the type of tester did have an effect on the weight loss.

Completely randomized design was an appropriate way to design this experiment. Factors that could be controlled, such as the thickness of the upholstery sample and type of abrasion paper, were easily controlled. Therefore the randomization of the samples is assumed to account for the factors that could not be controlled for. Even though this design is able to control for many factors, it is unable to answer all of the possible questions that researchers may have. The design and ANOVA were able to show that the type of abrasion tester did have effect on weight loss, however, it was not able to show which types had a greater effect.

Completely Randomized Design Code

To demonstrate completely randomized design in R, the variables “mpg” and “wt” from the dataset “mtcars” will be used. The variable “mpg” is the miles per gallon of fuel used and “wt” is the weight of the vehicle in thousands of pounds. For this example, randomness of the tests for mpg and weight are assumed. Based on this assumption, an ANOVA will be used to test if the weight of the vehicle as any effect on the gas mileage of the vehicle. A significance level of .05 will be used. The following R code is used:

>data(mtcars) #Before running a test, the dataset must be loaded

>dat<-mtcars

>#ANOVA_Format<-aov(response~factor, data=dataset)

>test_CRD<-aov(mpg~wt, data=dat)

>summary(test_CRD) #Display the results

The above R code uses the built-in ANOVA function to conduct the test. The code selects the selected dataset from the built-in sets in R and then conducts the test. Using the “Summary” function the following results are produced:

```
            Df Sum Sq Mean Sq F value   Pr(>F)    
wt           1  847.7   847.7   91.38 1.29e-10 ***
Residuals   30  278.3     9.3                     
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 2: Completely Randomized Design R output

As can be seen above, the output includes the degrees of freedom, sum of squares, mean squared error, F-value, and the p-value for the variable being investigated and the residuals. This ANOVA concludes that the weight of the car has an effect on the gas mileage of the car because the p-value of 1.29*10^-10 is less than the significance level of .05.

Randomized Block Design

Randomized Block Design reduces variability because the subjects or objects in the study are grouped into blocks based on their similar characteristics. In this type of design, the variability between the subjects within the same block is much less than that of the variability between the blocks. The subjects are divided into blocks and are then randomized within the blocks. This randomization helps to decrease the variability within the blocks, so that the variability between the blocks can be analyzed.

This type of design can be very useful because it reduces variability and allows the researcher to more clearly see the effect of a treatment on a group. Objects or subjects are not drawn from one sample, but instead from a block of similar objects or subjects. This reduces variability because objects are being drawn from levels with similar objects rather than being drawn from a randomized group.

Randomized Block Design Example

One important part of manufacturing is monitoring the impacts that the environment can have on the production process. This example deals with the effects that the environment can have on textile manufacturing. There are pests that can get into the cotton that is being manufactured and make it unusable for production. Zhang et al's 2010 study, "Effectiveness of thiamethoxam and imidacloprid seed treatments against Bemisia tabaci (Hemiptera: Aleyrodidae) on cotton" used randomized block design to investigate the effectiveness of different seed treatments against the pests.

The researchers used a randomized block design with three treatments and each treatment was replicated three times. The treatments included two different seed treatments and untreated seeds. The plots that were tested with each treatment were selected at random and then the effects of the pests on the plants were analyzed. In this experiment, the plants were divided into four blocks by age. The effects of each treatment on each age of plant were the analyzed. An ANOVA and corresponding F-test were used to analyze the outcomes of the experiments. This study attempted to control for error by randomly selecting the fields that would be used.

The randomized block design was the best design for this experiment because it was able to analyze whether or not the treatments had an effect on the amount of pests on the plants. The researchers were interested in multiple treatments, but not in the interaction between the treatments.

Randomized Block Design Code

The demonstration of randomized block design in R, will use the variables “mpag”, “wt”, and “cyl” from the “mtcar” dataset. It will analyze the effects of car weight on gas mileage by blocking on the number of cylinders that the car has. The following R code can be used to conduct the test:

```
>data(mtcars) #Before running a test, the dataset must be loaded
>dat<-mtcars
>#ANOVA_Format<-aov(response~factor + Blocking_Factor, data=dataset)
>test_RBD<-aov(mpg~wt + cyl,data=dat)
>summary(test_RBD) #Display the results
```

The code above produces the following results:

```
            Df Sum Sq Mean Sq F value   Pr(>F)    
wt           1  847.7   847.7  128.60 3.54e-12 ***
cyl          1   87.1    87.1   13.22  0.00106 ** 
Residuals   29  191.2     6.6                     
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 3: Randomized Block Design R output

Factorial Design

Factorial Design is used to study the response between factors and factor levels. In this type of study, every combination of the two are looked at. It investigates more than one independent variable. Factorial design uses an orthogonal model. This means that all of the independent variables need to be uncorrelated.

This type of design can be very time consuming. It is more useful when there are a smaller number of factors being researched, usually five or fewer.

Factorial Design Example

Kim et al. utilized factorial design in their 2003 study, "Sensitivity analysis for process parameters in GMA welding processes using a factorial design method." This study seeks to find and establish mathematical models that can be used to decide factors that improve the welding process. The factors that were investigated in the experiment were, arc current, welding voltage, and welding speed. These factors are especially important during automated welding, which can be very useful in manufacturing for saving time and money. The measured responses to these factors were bead width, bead height, and bead penetration.

The researchers repeated the experiment 27 times for each factor. They controlled for the type of steel plates, steel wire welding consumable, and weld conditions during the experiments. The welds were conducted in a lab that was able to be controlled for the appropriate data collection. The welds were then carried out by a robot. Once the tests were carried out, the data was collected by cutting open the welds to view the cross-section at the same point along each weld.

The response measurements were then analyzed using an ANOVA. The ANOVA showed that there was a statistically significant relationship between the welding factors and the measured responses.

A factorial design was an appropriate choice of design for this experiment because of the multiple factors and responses that the researchers were investigating. The other basic types of experimental design are unable to incorporate the level of complexity that this study needed.

Factorial Design Code

To demonstrate how to conduct a factorial design in R, the variables “mpg”, “wt”, and “cyl” will be used. The mileage of the car will be analyzed over all possible combinations of weight and number of cylinders. A .05 significance level will be used for this example. The following R code shows how to complete a factorial design:

```
>data(mtcars) #Before running a test, the dataset must be loaded
>dat<-mtcars
>#ANOVA_Format<-aov(response~factor*factor, data=dataset)
>test_FD<-aov(mpg~wt*cyl,data=dat)
>summary(test_FD) #Display the results
```

The above R code uses the built in ANOVA function. It measures the relationship between the mileage of the cars and all of the possible combinations between car weight and number of cylinders. The output shows the coefficients for each factor as well as for the interaction between the factors. Not all models will have the interaction terms. For some experiments, the interaction between the variables is not significant and therefore does not need to be included in the model. The significance codes in the R output indicate whether or not they are necessary. Also, an ANOVA with an F-test can be used to analyze the significance of the interaction terms in the model. It produces the following output:

```
            Df Sum Sq Mean Sq F value   Pr(>F)    
wt           1  847.7   847.7  151.21 8.35e-13 ***
cyl          1   87.1    87.1   15.54  0.00049 ***
wt:cyl       1   34.2    34.2    6.10  0.01988 *  
Residuals   28  157.0     5.6                     
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 4: Factorial Design R output

Analysis

The analysis of completely randomized, randomized block, and factorial designs is conducted using analysis of variance (ANOVA) or analysis of covariance (ANCOVA) calculations. The examples in R, in this thesis, all analyze the data through ANOVA or ANCOVA.

ANOVA is he simplest way to analyze the outcome of an experiment. It can be used to analyze one factor (one-way ANOVA) or two or more factors (two-way ANOVA). ANOVA focuses on the variation between the groups being tested. It investigates the effects of the factors on the outcomes. This test compares the means of different groups in the sample. ANOVA does not take into consideration the covariation between factors.

The ANCOVA calculation is much more complex that of the ANOVA. In an ANCOVA calculation, the covariate is used to reduce the experimental error. This test reduces the amount of error within the blocks, so that it can better test for differences between the blocks. ANCOVA reduces error by controlling variables that researchers are not interested in. These variables are called the covariate. By adding the covariate term to the calculation, the error is decreased.

Both ANOVA and ANCOVA are analyzed using an F-test. The F-test allows the researcher to determine at what level the results of the test are statistically significant. Both tests can be lengthy and tedious to calculate by hand. They can be easily run using the R software. The R code listed above can be used to demonstrate ANOVA and ANCOVA.

Selection of the Wrong Design

Choosing a too complex or a too simplistic design for an experiment can effect the outcome and analysis of the experiment. It is important to choose the proper design for the experiment in order to have the best results.

If a completely randomized block design is chosen, when a randomized block design or a factorial design should have been chosen, the error term of the experiment might be very high. This would be due to the multiple blocks or factors in the experiment that were not taken into account in the statistical analysis. By using a more complex design instead, the blocks or factors can be taken into account to reduce the error.

If a randomized block design or a factorial design is chosen, when the blocks or factors do not influence the outcome, the statistical significance of the outcome will be very low. This will show that there is little or no difference between the blocks or factors. Because there is little or no difference between the blocks, a completely randomized design could be used instead.

Discussion

This thesis only provided explanations and examples of three basic types of statistical experimental design. There are many more types and variations that can be used to investigate questions and problems. The techniques that are described in this thesis serve as building blocks for the more complex techniques for designing experiments.

Each technique attempts to control different factors and reduce error through blocking and the relationships between the block. Randomness is used to attempt to account for the factors that cannot be controlled. However, in every experiment there are unknown factors that cannot controlled. Some of these factors may not even be known. For example, they could include the impact of the air or water quality on the experiment. Figure 2 shows the process and the factors effecting the outcome of the experiment.

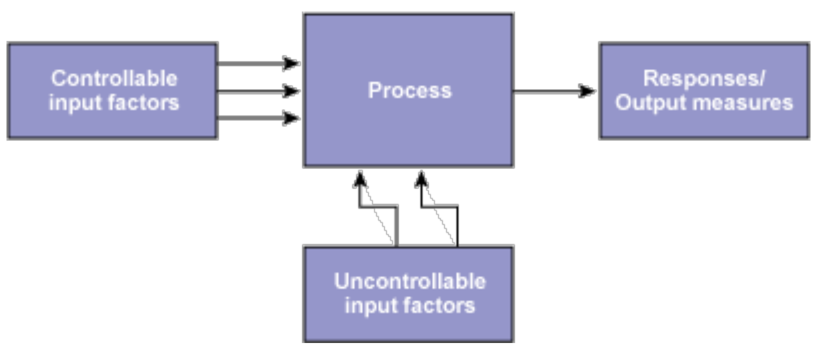

Figure 5: Impact of Factors

It can be very difficult to determine or control the uncontrollable and unknown factors. These factors can greatly change the outcome of the experiment. However, throughout the experiment, the design can be changed or adapted to account for the factors.

Completely randomized design is the easiest and most simple type of design to use, however, it also has the highest error rate.

Future Work

I hope to continue to learn about and research statistical design of experiments. After my graduation from the South Carolina Honors College, I will continue studying at the University of South Carolina to earn my Master of Science degree in statistics. During this time, I hope to conduct research and take courses that will further my statistical learning and interests. I have accepted an internship with the Boeing Company for the summer of 2019. The position is with the Applied Mathematics group that I previously mentioned. Working with this team will allow me to gain industry experience with many different statistical techniques.

After completion of my Master's degree, I will either continue my education or begin a career in industry as a statistician. No matter what path I decide to take, I hope to continue studying design of experiment techniques and the use of statistics in manufacturing and the aerospace industry.

Conclusion

In conclusion, there are many ways and variations that statistical studies can be conducted. The optimal way depends on the location, timing, and variables of the study. Experiments can be designed multiple different ways, each with varying statistical power and factors. Research and planning is necessary in order to conduct the most accurate experiment.

This thesis only described the basic techniques for experimental design. Completely randomized, randomized block, and factorial designs can be used for many experiments in a wide range of fields.

Bibliography

- https://www.britannica.com/biography/W-Edwards-Deming
- manufacturing. 2018. In *Merriam-Webster.com*. Retrieved November 6, 2018, from https://www.merriam-webster.com/dictionary/manufacturing
- Paur, Jason. "Boeing 787 Passes Incredible Wing Flex Test." *Wired*, Conde Nast, 4 June 2017, www.wired.com/2010/03/boeing-787-passes-incredible-wing-flex-test/.
- Jerkovic, Ivona & M Pallares, Josep & Capdevila, Xavier. (2010). Study of the abrasion resistance in the upholstery of automobile seats. Autex Research Journal. 10.
- Ribeiro, Jose Luis & Fogliatto, Flavio & Caten, Carla. (2000). Minimizing manufacturing and quality costs in multiresponse optimization. Quality Engineering. 13. 191-201. 10.1080/08982110108918641.
- Kim, InSoon & Son, K.J. & Yang, Yuansheng & Yaragada, P.K.D.V.. (2003). Sensitivity analysis for process parameters in GMA welding processes using a factorial design method. International Journal of Machine Tools and Manufacture. 43. 763-769. 10.1016/S0890-6955(03)00054-3.
- Zhang, Liping, et al. "Effectiveness of Thiamethoxam and Imidacloprid Seed Treatments against Bemisia Tabaci (Hemiptera: Aleyrodidae) on Cotton." *The Canadian Journal of Chemical Engineering*, Wiley-Blackwell, 12 Nov. 2010, onlinelibrary.wiley.com/doi/10.1002/ps.2056/abstract.
- Keppel, Geoffrey, and Thomas D. Wickens. *Design and Analysis: a Researchers Handbook*. Prentice Hall, 2003.