A SunCam online continuing education course

What Every Engineer Should Know About the Design and Analysis of Engineering Experiments I

by

O. Geoffrey Okogbaa, Ph.D., PE

Contents

Engineering Design

In any manufacturing or production activity, the most important concern is the ability to monitor the extent to which products and processes meet desired targets or specifications. Two culprits have been identified as major contributors or hindrances to product and process improvement and quality more than any other, namely, deviations from specifications; and excessive variability and noise around targets or specifications. Due to the complex nature of most production activity, these characteristics are part and parcel of the true state of nature in any production environment. As such, minimizing their effects requires detailed understanding and systematic analyses of their impacts and reach, using engineering and scientific design principles and techniques. In the developmental and conceptual design stages, designed experiments are utilized as tools to optimize these two quality characteristics in order to enhance process performance.

A good experimental design is one that furnishes needed information with minimum experimental effort. A good design requires the following:

a).The questions that are to be answered by the experiment must be clearly and carefully formulated.

b).Correct choice of experimental methods must be made, in light of the required accuracy and the various pitfalls likely to be encountered.

c).The general pattern of experiment (sample size, spacing, the interrelationship of the observations) must be clearly identified.

In general, experiments are designed in an attempt to reduce the noise and increase the volume of the signal, in order to maximize the quality of information obtained at a reasonable and fixed cost.

There are several fundamental issues that are central to the successful design of experiments for process and quality improvements. The ability to successfully design, conduct, and interpret planned experiments for product and process design and improvement depends on how well these elements are understood and operationalized. An approach to understanding this process is encapsulated in the strategy often referred to as the DMAIC (Design-Measure-Analyze-Improve-Control) process.

Fundamentally, DMAIC is a data-driven quality strategy for improving processes and is an integral part of a company's Six Sigma Quality Initiative. Each step in the cyclical DMAIC process is necessary to ensure the best results. The DMAIC process steps are listed as follows:

- **Define** the Customer, their Critical to Quality (CTQ) issues, and the Core Business Process involved. Define who customers are, their requirements for products and services, and their expectations. Define project boundaries as the stop and start of the process. Define the process to be improved by mapping the process flow.

- **Measure** the performance of the Core Business Process involved. Develop a data collection plan for the process. Collect data from many sources to determine types of defects and metrics. Compare to customer survey results to determine shortfall.

- **Analyze** the data collected and process map to determine root causes of defects and opportunities for improvement. Identify gaps between current performance and goal performance. Prioritize opportunities to improve. Identify sources of variation.

- **Improve** the target process by designing creative solutions to fix and prevent problems. Create innovate solutions using technology and discipline. Develop and deploy implementation plan.

- **Control** the improvements to keep the process on the new course. Prevent reverting back to the "old way". Require the development, documentation and implementation of ongoing monitoring plan. Institutionalize the improvements through the modification of systems and structures (staffing, training, incentives)

Role of Experiments in the Engineering Design Process

A serious shortcoming of past approaches to process performance has been the inability to deal rationally with the performance issue early in the product and process development life cycle. Over the past several decades, it has become clear, through the work of Taguchi and others, that parameter selection during the early stages of product and process design can be improved by measuring performance as it relates to functional variation during use and by proper design of experiments. In particular, the concept of robust design, advocated by Taguchi as part of his model for the design process, has proven to be an effective tool for product and process design and improvement (see figure 1.)

There is an important distinction to be made between testing and experimentation. While both have their rightful place, one should not serve as an alternative for the other. The Japanese have used experimentation, namely "design of experiments", for parameter selection at the product and process design stages. In this case, the objective is to experiment with various combinations of the important design parameters for the purpose of identifying the particular combination(s) that optimize certain design criteria or performance metrics.

In the past, a great deal of emphasis was placed on life testing by subjecting many identical units to field conditions in order to determine the life expectancy with regard to performance. When a design fails, changes are made and the product is retested. The question naturally is: How are those changes identified? The answer, most often, is to adopt a deliberate experimental approach or do more ad hoc procedures. Thus, the life testing of product performance is important, but not a

substitute, for experimentation to determine product and process capability or performance in the long or short term.

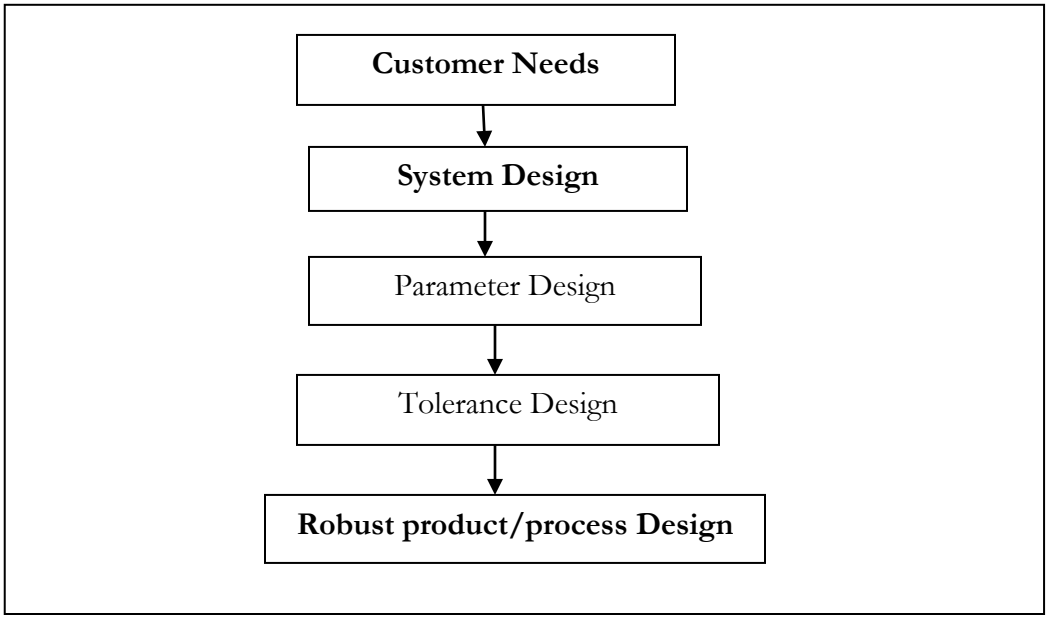


Figure 1 Taguchi's Method for the Engineering Design Process

The Role of Statistics and Probability in Engineering Design

3.1 Review of Statistical Inference

In many engineering settings, there are typically large numbers of random quantities or random variables. More often than not, we do not know the probability structure of these variables or their underlying characteristics. However, we do want to determine these quantities in order to have better control of the system operation, which we can accomplish by taking observations on the variables. But, we cannot take readings carelessly because there are biases, errors, and noise inherent in any such process. Based on the classical definition of probability, the determination of the probability or the expected value associated with the random variables would require an infinite number of observations. However, having only samples of finite sizes, we can usually estimate the values in question in the form of sample statistics.

The final result of a statistical inference is always a decision to act or not to act. In some instances, the decision could be to accept the observed or computed value of the estimator in place of the unknown parameter, without requiring that it be exactly the true value. On the other hand, we may decide to reject or not reject the assumptions about a certain distribution without conceding that such a statement is true beyond doubt. The use of statistical inference enables us to control the

errors that could arise as a result of our decisions and to ensure that these errors, while inevitable, are as small as economically possible.

Inferential statistics has two main branches, namely estimation and test of hypotheses. For a good estimation, a fairly large sample is needed. In some cases, only very limited samples may be available. Such limitation means that the distribution is already known or assumed before-hand and thus, the ensuing analysis verifies that the distribution has not changed. Estimation and tests of hypotheses are avenues to substantiate such assumptions or claims.

3.2 Estimation

There are two types of estimators (Figure 2), namely point estimators and interval estimators. Two general methods generate estimators of parameters, namely, the methods of moments and maximum likelihood. For some problems, both the method of moments and maximum likelihood lead to exactly the same estimators, and for others they do not. When the two methods do not agree, the maximum likelihood estimator is usually preferred.

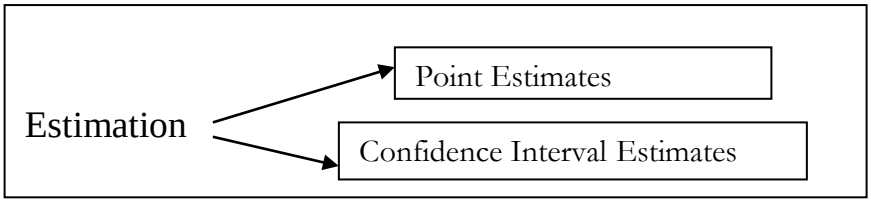


Figure 2 The two types of Estimators

3.3 Point Estimates

A point estimate is a single value or point on the real line, which we feel is a good guess for the unknown population parameter value that is being sought. The motivation for conducting an experiment stems from the understanding that, in most cases, it is impractical to obtain the value of the parameter that we seek, since it would require the almost impossible task of observing the outcome of an infinite population. This being the case, the problem then reduces to one of attempting to extract as much information as possible about the parameter from the sample(s) based on the sample statistic. In other words, point estimates are summary statistics that capture the essence of the sought-after parameter.. However, there are several ways to represent this information.

As an example, in estimating the central tendency, which is a population parameter, it is generally agreed that the mean and the median are both reasonable quantities with which to measure such a parameter. Also, in estimating the variance of a random variable, the sample variance and the range are both used as estimators. Obviously, only one of these estimates can be used or employed at any one time. Thus, there needs to be a set of criteria, standards, or properties by which to judge or characterize the estimators. The properties of unbiasedness and efficiency are some of the properties that are desired in a good estimator. A statistic X is called 'best unbiased estimator

(BUE)' for the parameter θ, if the statistic is unbiased and efficient, i.e., if E (X) = θ and if the variance of X is less than or equal to the variance of every other unbiased statistic.

Single or Point Estimates of the mean:

$$\left.\begin{cases} \overline{X} & = Mean \\ \tilde{X} & = Median \\ M_0 & = Mode \\ (X_L + X_S)/2 & = Mid\ Range \end{cases}\right\} \ all\ \ are\ unbiased\ \ estimates$$

In the case of the sample mean and median as estimators of the population central tendency, both are unbiased estimators, i.e., E(X) = θ and $E(\tilde{X}) = \theta$. However, the best unbiased estimator (BUE) is the sample mean because it has the minimum variance with respect to all the estimators of θ. In the case of the sample mean and median as estimators of the population central tendency, both are unbiased estimators, i.e., E ($\overline{X}$) = θ and $E(\tilde{X}) = \theta$. The variance of the sample mean and that of the median are as shown.

For the Mean, $V(\overline{X}) = \sigma_{\overline{X}}^2 = \frac{\sigma^2}{n}$, For the Median, $V(\tilde{X}) = \sigma_{\tilde{X}}^2 = \frac{1.57\sigma^2}{n}$

The variance of the median is 1.57 times the variance of the sample mean. Therefore, using the criteria for BUE, the sample mean is considered the BUE because it has the minimum variance with respect to all the estimators of θ. As noted previously, both the Mid-range and the Mode are also unbiased estimators of the population mean but they are not BUE.

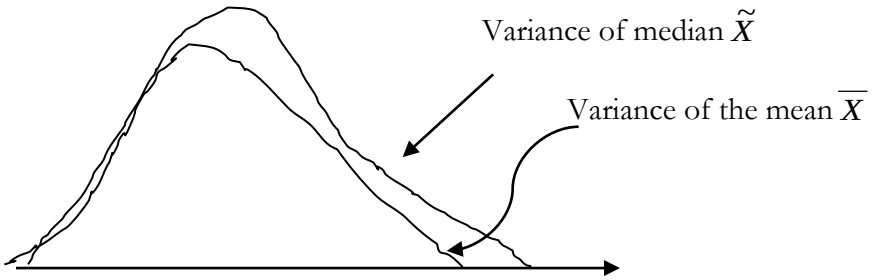


Figure 3 Variance of Mean and Median

3.3.1 Point Estimates for the Mean and Variance of the Population

The following are the point estimates for the mean and variance. For the mean, we have

$$\mu_x = \frac{\sum X}{n}, \ \ and \ \ \ \mu_{\bar{x}} = \frac{\sum_{i=1}^{k} \mu_x}{k} \ \ \ \sigma_{\bar{x}}^2 = \frac{\sigma_x^2}{\sqrt{n}}, \ where: \sigma_{\bar{x}}^2 = \frac{\sum (\overline{X} - \overline{\overline{X}})}{k-1}$$

Where *k* =number of subgroups and n is the sample size

3.3.2 Central Limit Theorem

The central limit theorem (CLT) is a statistical theory that states that, given a sufficiently large sample size from a population with a finite variance, the mean of all samples from that population would be approximately equal to the mean of the population. Let

$$\overline{X}_1, \overline{X}_2, \overline{X}_3, \ldots\ldots \overline{X}_n$$

denote the measurements or output of a random sample of size n from any distribution having finite variance σ^2 and mean μ, then the random variable:

$$\frac{\sqrt{n}\left(\overline{X}-\mu\right)}{\sigma_X}$$

has a limiting normal distribution with zero mean and variance equal to unity. In other words, even though the individual measurements have a distribution that is not the normal distribution, the distribution of the sample means

$$\overline{X}_1, \overline{X}_2, \overline{X}_3, \ldots\ldots \overline{X}_n \text{ as } n \to \infty$$

tends to be approximately normally distributed. In other words, the sampling distribution of the sample means is the normal distribution When this condition is true, it would be possible to use this property to compute approximate probabilities concerning the distribution and to find an approximate confidence interval for μ and to test certain hypotheses without knowing the exact distribution of μ in every case or situation.

The CLT establishes that, for the most commonly studied scenarios, when independent random variables are added, their sum tends toward a normal distribution even if the original variables themselves are not normally distributed. This is very important, especially because it is often difficult to determine the underlying parent distribution, which is needed to determine the probabilities of event occurrence to enable engineering decisions to be made in an informed manner.

3.3.3 Sampling Distribution for the Mean

The sampling distribution of the sample mean is the normal distribution based on the CLT. In other words, the distribution of the sample mean $\overline{X}$ is the normal distribution with mean:

$$\mu_{\overline{X}} = \overline{\overline{X}}, \text{ and } \sigma^2_{\overline{X}} = \frac{\sigma^2}{n}$$

A sample of 25 spur gears is taken from the lot resulting in the Diametral-Pitch **(**DP**)** measurements are as shown in table 1. Another sample could be taken from the lot with a mean value: $\overline{X}_1 = 4\,inches.$ yet another sample from the same lot could yield a DP of: $\overline{X}_1 = 6\,inches.$ The larger the DP, the higher the stress on the gear tooth. Let the average, the DP = 4 inches with S= 0.55 inches. Find the probability that some of the gear-spurs will not meet the requirement, that is: $P(\overline{X} > \mu_0)$.

Table 1: Diametral Pitch (DP) Measurements (in.)

S/N	Diametral Pitch (DP inch)	S/N	Diametral Pitch (DP inch)
1	2.442966	14	4.616154
2	5.870707	15	3.914669
3	4.127012	16	5.484784
4	2.060597	17	3.387145
5	5.96805	18	2.976296
6	3.022355	19	4.135116
7	3.301695	20	4.422885
8	5.172247	21	5.738772
9	5.341773	22	4.407002
10	4.402271	23	4.690581
11	2.806244	24	3.997262
12	4.831229	25	3.40529
13	3.224622	**Mean**	4.149909
		Std. Dev	1.09433

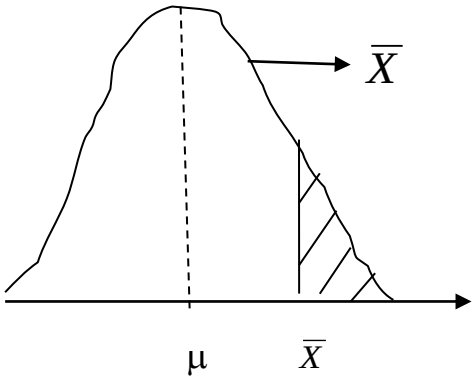


Figure 4 Sketch of Probability Distribution for X-bar

$$\textit{Given: } \mu_{\overline{X}} = 4, \;\; \sigma_{\overline{X}} = \frac{0.55}{\sqrt{25}} = 0.11$$

$$P(\overline{X} > \mu_0) = P\left(\overline{X} > \frac{\overline{X_0} - \mu_0}{\sigma_{\overline{X}}}\right) = 1 - \Phi\left(\frac{4.149909 - 4.0}{0.11}\right) = 1 - \Phi(1.363)$$

$$\Phi(1.363) = 0.9136, \; \textit{Shaded area} = (1 - 0.9136) = 0.0864 \cong 9\%$$

There is only a 9% chance that the spur-gears from that population will not meet the requirement. Note that we did not use the standard deviation we computed for the data for the problem. Why? You will recall that the focus is on the sampling distribution for the mean. The mean is the random variable in this case. Later, we will consider the sampling distribution for the variance based on the variance from the data. Each sample we take, just like the mean, will yield a variance statistic, since the variance is a random variable. **Due to the unbiased nature of the sample mean as an estimator of the population mean, the sampling distribution of two or more means is normally distributed. The sum of the means is also normally distributed**.

3.1.4 Sampling Distribution for the Mean when sample size n<30

Student-*t* distribution is similar to the standard normal distribution when the sample size is small, typically $n<30$. Characteristics of Student-*t* distribution include the following:

1).The probability distribution appears to be symmetric about $t = 0$ just like the standard normal distribution

2).The probability distribution appears to be bell-shaped.

3).The density curve looks like a standard normal curve, but the tails of the *t*-distribution are "heavier" than the tails of the normal distribution. That is, we are more likely to get extreme *t*-values than extreme z values. The nice thing about the *t* -distribution, is that we can use it when the sample size does not justify the use of the standard normal, that is, when $n<30$.

Recall that in the case of the Standard Normal Variable, the random deviate:

$$Z = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}}$$

In the case of the t distribution, the random variable, *t,* is give as:

$$t = \frac{\overline{X} - \mu}{S / \sqrt{n}}, \text{ with (n-1) DF}$$

3.3.4 Sampling Distribution for the Sample Variance

From the CLT, we know that the distribution of the sample mean is approximately normal. Unfortunately, unlike the sample mean, there is no CLT analog for the variance. However, when the individual observation X_is are from a normal distribution, there is a special condition under which we can consider the sampling distribution of the sample variance. Suppose, as indicated earlier, X_1, X_2, .. , X_n are from a normal distribution, N (μ,σ^2), and we will recall that the CLT applies to any arbitrary distributions. If this is true, the distribution of the sample variance is related to the Chi-Square distribution For the $X_1, X_2, \ldots, X_n$, $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is the mean, and $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X - \overline{X}\right)^2$ is the sample variance, then $\frac{(n-1)S^2}{\sigma^2}$ is the chi-square distribution with (n-1) degrees of freedom. The table for the Chi-Square is available in most basic statistics texts.

3.3.5 Sampling Distribution for Two Variances

When the variances from two or more sources are being evaluated, the resulting sampling distribution follows the Snedecor's F-distribution or simply the F-Distribution. The sampling distribution for two variances is used to test whether the variances of two populations are equal. The F distribution is given as: $F = \frac{S_1^2 / \sigma_1^2}{S_2^2 / \sigma_2^2}$, with (υ_1, and υ_2), where $\upsilon_1 = n_1 - 1$ and $\upsilon_2 = n_2 - 1$;

where the notation of 1 or 2 is perfunctory and depends on which variance is larger.

Please note, that for ease of computation, it is recommended that when taking ratios of sample variances, we should put the larger variance in the numerator and the smaller variance in the denominator. We will see how this is done with a numerical example later. In order to use this test, the following must hold:

- Both populations are normally distributed
- Both samples are drawn independently from each other.
- Within each sample, the observations are sampled randomly and independently of each other.

3.4 Interval (Confidence interval) Estimators

In practical situations, there are two types of estimation problems. In one case, we may have a constant, ϕ, a theoretical quantity that must be determined by means of measurements. Some examples include: the time it takes to complete a machining operation, the amount of yield from a given reaction, the number of material handling moves required for a certain material handling type, and so on. The result, Y, of the measurement activity is a random variable whose distribution function depends on the constant, ϕ, and perhaps other quantities.

If we want to use a single number in place of the unknown constant or parameter, then point estimation is the appropriate method. If we are using a good estimator (unbiased and efficient), then the resulting estimate should be close to the unknown true value. However, we know that the estimator is subject to error of measurement (in the case of the constant) and variability (in the case of the random variable). Consequently, it is instructive to have some information on the deviation from the true value (in our case, the true mean or the true deviation). This is where confidence intervals come in. Due to the variability or error in measurement, we want to establish an interval within which we would expect to find the value we seek. In other words, in repeated sampling and using the same method for selection , we would expect the true parameter value to fall within a specified interval a given percent of the time. For example, a 95% confidence interval means that in repeated sampling and using the same sampling method, we would expect the true parameter value to fall within our confidence interval 95% of the time. Let us do some housekeeping before we delve deep into the area of Confidence Intervals. First, let us look at the error associated with the estimate $\overline{X}$.

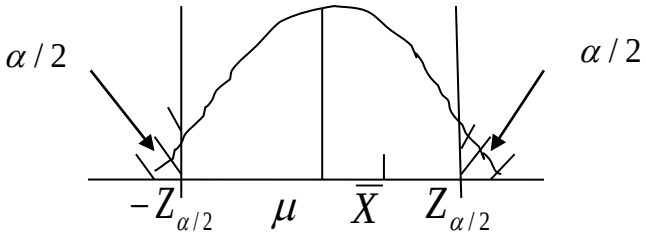


Figure 5 Confidence Interval for the Mean

We know that $Z = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}}$

is true for either positive or negative depending on where Z is relative to the mean μ.

$$-Z_{\alpha/2} < \frac{\overline{X} - \mu}{\sigma / \sqrt{n}} < Z_{\alpha/2}$$

Reworking, we have the confidence interval for the mean which is a probability statement given by:

$$P\left(\overline{X} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$$

3.4.1 Determination of Sample Size

If we examine the error associated with the mean X-bar, say E, where E is given by: $E = \overline{X} - \mu$, we can re-express Z as follows

$$-Z_{\alpha/2} < \frac{E}{\sigma/\sqrt{n}} < Z_{\alpha/2}, \Rightarrow E = \pm\left(Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right) \Rightarrow |E| = \left(Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right)$$

$$n = \left[\frac{Z_{\alpha/2}\sigma}{E}\right]^2 \text{ if } \sigma \text{ is known}$$

$$n = \left[\frac{t_{\alpha/2}S}{E}\right]^2, \text{ if } \sigma \text{ is unknown but estimated from S}$$

You might ask what this all means or why we need a value for *n*. The problem is that ordinarily the value of n is not given; thus, the sample size is unknown. Customarily, a company may have a policy dictating the size of the process errors that has been determined using historical data. The given value and the level of confidence specified based on the data will help determine the sample size needed to cover the error. For example, a company may say that it is comfortable with an error of ±10%, this being the error between the true mean and the estimated mean.

Example: A company is willing to accept an error of ± 15% with 90% confidence. a). Assuming that the variance is known and σ =1.5 units. What sample size will guarantee this level of protection? b) Assuming that variance is unknown and that somehow the company has an estimated value from experience for the process sample standard deviation, S=2 with 90% confidence. What sample size will be required? $n = \left[\frac{Z_{\alpha/2}\sigma}{E}\right]^2$ if σ is known

α = (1-0.90) = 0.1, $\alpha/2$=0.05, E=0.15. from the standard normal table $Z_{0.05}=Z_{0.95}=1.645$
Strictly speaking, there is no way we can evaluate this without knowing the sample size. Remember that to evaluate the t-statistic we need the degrees of freedom equal to n-1. Even though the variance is unknown, we do have the estimate of S determined historically. Therefore, we can use the *Z* distribution in place of the t-distribution to evaluate the sample size. Note that the *t*-statistic and the *Z*-statistic are identical when *n* = infinity. In this case, use the value of *t*-statistic with $\upsilon = \infty$.

$$n = \left[\frac{t_{\alpha/2}S}{E}\right]^2 = \left[\frac{(1.645)2}{0.15}\right]^2 = 480$$

3.4.2 One-Sided Confidence Intervals

Under certain conditions, only one-sided intervals may be of interest. For example, in the case of still bars, we want the measured strength to be as high as possible. Therefore, our main concern is that the strength values do not go beyond a certain lower limit. Accordingly, we would establish a lower confidence (one-sided) interval. By contrast, we may have a variable (the number of defects) whose value we would want to be as close to zero as possible. For this, we are only concerned with how high the value can go, so we establish a one-sided confidence interval. A one-sided confidence interval is looked at as a one-tailed interval (upper confidence level (UCL) or lower confidence level (LCL), but not both) unlike the two tails of the two-sided confidence interval. That being the case, we use α rather than $\alpha/2$.

$$UCL_t:\ P\left(\mu < \overline{X} + t_{\alpha,\upsilon}\frac{S}{\sqrt{n}}\right) = 1-\alpha, \qquad LCL_t:\ P\left(\mu > \overline{X} - t_{\alpha,\upsilon}\frac{S}{\sqrt{n}}\right) = 1-\alpha$$

$$UCL_Z: P\left(\mu < \overline{X} + Z_{\alpha,}\frac{(S\ or\ \sigma)}{\sqrt{n}}\right) = 1-\alpha, \qquad LCL_Z:\ P\left(\mu > \overline{X} - Z_{\alpha}\frac{(S\ or\ \sigma)}{\sqrt{n}}\right) = 1-\alpha$$

$$UCL_\sigma: P\left(\sigma^2 < \frac{24(4.9)^2}{\chi^2_{0.975,24}}\right) = 1-\alpha, \qquad LCL_\sigma: P\left(\sigma^2 > \frac{24(4.9)^2}{\chi^2_{0.025,24}}\right) = 1-\alpha$$

Let us use an example to illustrate. For a product C and n=25, let the mean of the grinding duration $\overline{X}$ = 15 minutes, S =1.5 minutes.. Find the 95% lower confidence interval (LCL) for the of product C. LCL_t. Since n<30, we will assume that the sampling distribution is the Student-t distribution with the sample statistic equal to the t

$$P\left(\mu < \overline{X} - t_{\alpha,\upsilon}\frac{S}{\sqrt{n}}\right) = 1-\alpha$$

$$[1.761(1.5)]/5 = 0.5283,\ LCL = 15.0 - 0.5283 = 14.358$$

$$P(\mu < 14.358) = 95\% \quad \left(t_{0.05,14} = 1.761,\ from\ the\ t-table\right)$$

3.5 Test of Hypothesis

A test of hypothesis is a test on an assumption or statement that may or may not be true concerning the parameter of the population of interest. An examination of the entire population determines the truth or falsity of such a test. Since this is impractical in most situations, a random sample is taken from the population and the information used to deduce whether the hypothesis is

likely true or not. Evidence from the sample that is inconsistent with the stated hypothesis leads to a rejection, whereas evidence supporting the hypothesis leads to its acceptance. The acceptance of a statistical hypothesis does not necessarily imply that it is true**.** Thus, hypotheses that are formulated with the hope of rejecting are called null hypotheses and denoted by H_o. The rejection of H_o leads to the acceptance of an alternate hypothesis denoted by H_1. The decision to reject or not reject a hypothesis is based on the value of the test statistic. The test statistic is compared to a critical value. The critical value is based on the level of significance of the test and represents values in the critical region, as defined by the significance level.

3.5.1 Errors Associated With Decisions on Test of Hypothesis

A decision to reject or not reject a test leads to two types of errors. The error is the result of a decision made based on information from a sample rather than the actual process population itself. The fact is that we are trying to ascertain the true state of nature using information from the sample. We, of course, do not know the true state of nature and would like to infer it from the sample. This notion is perhaps one of the most important foundations of statistics. So, while we do seek the population value, we can only approach it by way of the sample value, which is of limited value unless it points us to or gives us the population value.

All samples are taken, not for their own sake, but to provide information or inference about the population value. There are two types of errors that are possible in statistical testing: **Type I (α)**, and **Type II (β)**. A type I error is committed when the null Hypothesis (H_0) is rejected. Alternatively, a type II error is committed when the null Hypothesis (H_1) is not rejected. This is loosely referred to as accepting the null Hypothesis. These errors are aptly demonstrated by the schematic in Table 2.

		TRUE STATE OF NATURE	
DECISION	**Accept**	H_0 **True** NO ERROR	H_0 **False** TYPE II ERROR
DECISION	**Do not Accept**	TYPE I ERROR	NO ERROR

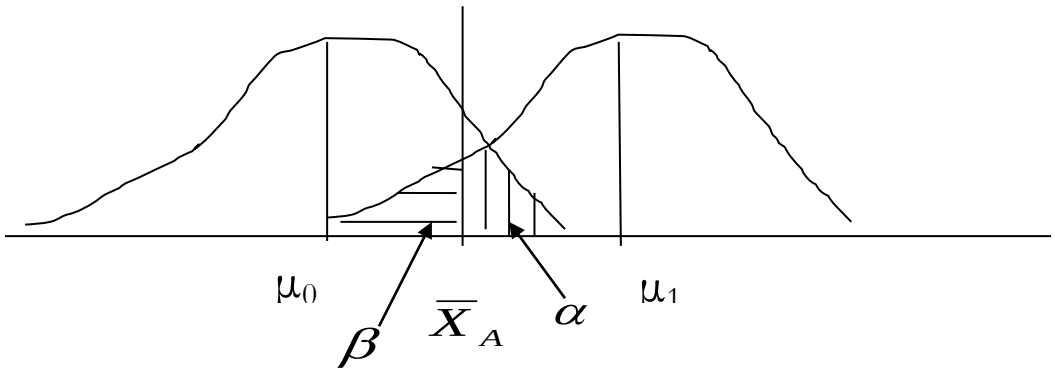
Table 2 Schematic for Type I and Type II Errors

3.5.2 Computation of the Required Sample Size (n) given (α) and (β)

In order to exert some control over a process, the engineer might specify the size of both Type I and Type II errors that the system can tolerate. The next step is to determine what value of n (sample) would help guarantee the level of protection based on these error levels. When the underlying process is normally distributed or when our focus is on the mean of the process, as you

may recall, even if the process is not normally distribution according to the central limit theory, the means from the process follow the normal distribution. Assume we have specified α and μ_0. If we also specify β, then we must necessarily specify μ_1. A sketch of the relationship between these parameters will help explain the procedure (figure 6).

Figure 6 Location of α and β



$$\frac{\overline{X_A}-\mu_1}{\sigma/\sqrt{n}}=-Z_\beta \Rightarrow\Rightarrow \overline{X_A}=\mu_1-Z_\beta\left(\sigma/\sqrt{n}\right).....(1)$$

$$\frac{\overline{X_A}-\mu_0}{\sigma/\sqrt{n}}=Z_\alpha \Rightarrow\Rightarrow \overline{X_A}=\mu_0+Z_\alpha\left(\sigma/\sqrt{n}\right).....(2)$$

$$Solving:\ (1)-(2):(\mu_1-\mu_0)-Z_\beta\frac{\sigma}{\sqrt{n}}-Z_\alpha\frac{\sigma}{\sqrt{n}}=0\Rightarrow(\mu_1-\mu_0)=\frac{\sigma}{\sqrt{n}}(Z_\alpha+Z_\beta)$$

$$\frac{1}{(\mu_1-\mu_0)}=\frac{\sqrt{n}}{(Z_\alpha+Z_\beta)\sigma}\Rightarrow n=\frac{(Z_\alpha+Z_\beta)^2}{\left(\frac{\mu_1-\mu_0}{\sigma}\right)^2}=\frac{(Z_\alpha+Z_\beta)^2}{\Delta^2},\ where:\Delta=\frac{(\mu_1-\mu_0)}{\sigma}$$

Let: μ_0=100, σ=10, α=0.05. Let β=0.1 for μ_1=110.

Compute n that will provide the level of protection given by the Type I and Type II errors.

$$n=\frac{(Z_{0.05}+Z_{0.10})^2}{\left(\frac{110-100}{10}\right)^2}=\left(\frac{1.645-1.282}{1}\right)^2=(2.927)^2=8.57\approx 9$$

For $\mu_1>\mu_0$, $Z_\beta=\Delta\sqrt{n}-Z_\alpha$, For $\mu_1<\mu_0$, $Z_\beta=1-\Phi\left(-Z_\alpha+\sqrt{n}\Delta\right)$

$$\text{For } \mu_1\neq\mu_0,\ Z_\beta=\Delta\sqrt{n}-Z_{\alpha/2},\ n=\frac{(Z_\alpha+Z_\beta)^2}{\left(\frac{\mu_1-\mu_0}{\sigma}\right)^2}$$

$$\text{for both } \mu_1>\mu_0,\ \mu_1<\mu_0, \text{ and: } n=\frac{(Z_{\alpha/2}+Z_\beta)^2}{\left(\frac{\mu_1-\mu_0}{\sigma}\right)^2}\quad \text{for } \mu_1\neq\mu_0$$

3.5.3 Hypotheses Test for one mean

1.Specify the null and the alternative hypothesis.

H_o: μ_0 = 100 g, H_1: μ_0 > 100 g

2. Specify the data needed and the sample size for the test for this one-sided test**.**

Note that since σ is known or given, then the Test Statistic is the Z (normal distribution).

α = 0.05, Hence $Z_{0.05}$=$Z_{0.95}$ =1.645, n =9, σ = 10, $\overline{X}$ =103

3. Critical region: Reject if $Z > Za$

4. Compute the Test Statistic,

$$Z = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{(103-100)}{\frac{10}{3}} = \frac{(103-100)(3)}{10} = 0.9$$

5**. Decision: Since $Z < Z\alpha$ (0.9<1.645), Therefore cannot Reject null (H_0).**

Question: How high a value of the Sample can we get before we reject the null hypothesis?

$$\frac{(\overline{X}-100)(3)}{10} = 1.645 \Rightarrow \frac{1.645(10)}{3} = \overline{X} - 100 \Rightarrow \overline{X} = 100 + 5.5 = 105.5$$

Thus, if we get a value of $\overline{X} \geq 105.5$, then we will reject H_0.

3.5.4 Hypothesis Test for One Variance

1.Specify the null and the alternative hypothesis.

H_o :σ^2 = 40 , H_1 :σ^2: > 40

2. Specify the data needed and the sample size for the test for this one-sided test.

α = 0.01, n =12, σ^2 = 40, S^2=62. Test Statistic is the Chi-Square.

3. Compute the Test Statistic.

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2} = \frac{(11)67}{40} = 17.05$$

4. Critical region: reject if $\chi^2 > \chi^2_{0.01,11}$ $\chi^2_{0.01,\ 11} = 24.725$

5. Decision: Since $\chi^2 < \chi^2_{0.01,11}$ (17.05<24.725), Therefore cannot Reject null (H_0).

3.5.5 Hypothesis Test for Two Variances

1. Specify the null and the alternative hypothesis.

$$H_0 : \sigma_1^2 = \sigma_2^2,\ H_1 : \sigma_1^2 \neq \sigma_2^2$$

2. Specify the data needed and the sample size for the test for this one-sided test**.**

$$\alpha = 0.01,\ n_1 = 8, \overline{X}_1 = 42.2\,g,\ S_1^2 = 10\,g^2, n_2 = 15, \overline{X}_2 = 44.5\,g,\ S_2^2 = 18\,g^2$$

3. Compute the Test Statistic is the F-distribution:

$$F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} = \frac{S_1^2}{S_2^2} = 10/18$$

4. Critical region: reject if $F > F_{0.05,7,14}$ *or* $F < F_{0.95,7,14}$, $v_1 = n_1 - 1 = 7$, $v_2 = n_2 - 1 = 14$,

Note that $F_{\alpha/2,\,\upsilon_1,\upsilon_2} > F_{(1-\alpha/2),\,\upsilon_1,\upsilon_2}$, **Also:** $F_{1-\alpha,\,\upsilon_1,\upsilon_2} = \dfrac{1}{F_{\alpha,\,\upsilon_2,\,,\upsilon_1}}$

$$F_{0.05,7,14} = 2.76, \quad F_{0,95,7,14} = \frac{1}{F_{0.05,14,7}} = \frac{1}{3.53}$$

5. Decision: Since $F > F_{0.95,7,14}$ $\left(10/18 > 1/3.53\right)$**, Therefore cannot Reject null (*H₀*)**

Since we cannot reject H_0, we can assume that the two variances are equal. This fact determines what type of pooled variances we can use to test the hypothesis of two means in this problem. Given this test result, we will use the pooled variance as follows: $S_p^2 = \dfrac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1 + n_2 - 2}$

3.5.6 Hypotheses Test for Two Means

1. Specify the null and the alternative hypothesis.

$H_0 : \mu_1 = \mu_2$, $H_1 : \mu_1 \neq \mu_2$

2. Specify the data needed and the sample size for the test for this one-sided test**.**

Note that since σ is an unknown equal and the sample sizes are <30, then the Test Statistic is the Student-t distribution.

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{S_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}, \; where \; (\mu_1 - \mu_2 = 0)$$

and: $S_p^2 = \dfrac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1 + n_2 - 2} = \dfrac{(7)(10) + (14)(18)}{21} = \dfrac{10 + 36}{3} = 15.33$

$\alpha = 0.05,\; n_1 = 8, \bar{X}_1 = 42.2\,g,\; S_1^2 = 10\,g^2, n_2 = 15, \bar{X}_2 = 44.5\,g,\; S_2^2 = 18\,g^2, df = \upsilon = 21$

3. Critical region: Reject if $|t| > t_{\alpha/2,\upsilon} \Rightarrow |t| > t_{0.025,21}, but: t_{0.025,21} = 2.08$

4. Compute the Test Statistic,

$$t = \frac{2.3}{\sqrt{15.33\left(\frac{1}{8} + \frac{1}{15}\right)}} = 1.34$$

5. Decision: Since $|t| < t_{0.025,21}, (1.34 < 2.08)$ **Therefore cannot Reject null (*H₀*).**

3.5.7 Operating Characteristic Curve

Operating characteristic curves are useful tools for exploring the power of a control process. OC curves provide a mechanism to gauge how likely it is that a sample statistic is not outside of the control limits when, in fact, it has shifted by a certain amount. This probability is usually referred to as β or Type II error probability, which is the probability of erroneously accepting the 'true state of

nature' (e.g. mean, variance, etc.) as being a given value, when in fact it is not. Note that operating characteristic curves pertain to the false-acceptance probability. The sample size for establishing an OC curve is determined by the cost of implementing the plan (e.g., cost per item sampled) and on the costs resulting from *not* detecting quality problems and thus passing unfit products.

3.5.8 Computation of the Parameters of the OC Curve

A Type II error (β) is the probability of accepting the original hypothesis H_0 when it is not true or when some alternative hypothesis, H_1 is true. Thus, β is a function of the value of the test statistic that is less (or greater) than the hypothesized value. Suppose we start with the hypothesis: $H_0 : \mu_0 = 12,\ H_1 : \mu_0 > 12$. Let n=16, and σ=7. α=0.05, Z_α=1.645

Recall: $Z_\beta = \Delta\sqrt{n} - Z_\alpha$, *where* $\Delta = \dfrac{(\mu_1 - \mu_0)}{\sigma}$.

We can now examine how β varies.

For μ_1=13, Z_β=(1/7)(4) -1.645 = -1.074

β =1-ϕ(-1.074)=1- {1-ϕ(1.074)}=1-0.1

For μ_1=15, Z_β=(3/7)(4) -1.645 = 0.0692

β=1-ϕ(0.074)=1-0.53=0.47

For μ_1=17, Z_β= (5/7)(4) -1.645 = 1.212

β=1-ϕ(1.212)=1-0.8869=0.11

Table 3 Computation of the parameters of the OC Curve

μ_1	Δ	$Z_\beta = \Delta\sqrt{n} - Z_\alpha$	β	**1-β (power of the test)**
13	1/7	-1.074	1-0.14= 0.86	0.14
15	3/7	0.0692	0.47	0.53
17	5/7	1.212	0.11	0.89
19	1	2.355	0.01	0.990
21	9/7	3.50	0.001	0.999
26	2	6.33	0	1

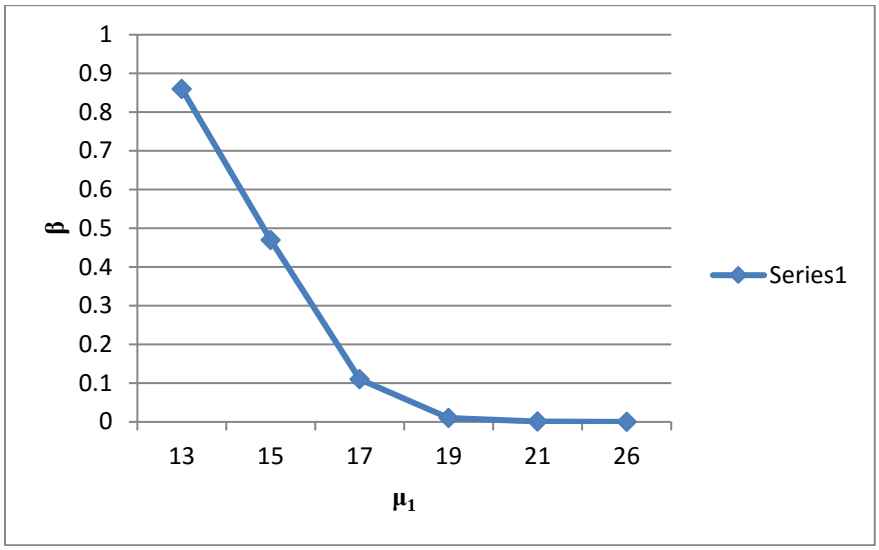


Figure 7 Plot of the Operating Characteristic Curve

Purpose and Nature of Planned Experiments

Scientists and engineers are involved in experimentation as a means to describe, predict, and control any phenomena of interest. The collection of data is a fundamental activity for the building and verification of mathematical models, whether such models are derived from first scientific principles or are purely empirical in nature. Comparative experiments are an important means to discern differences in the behavior of processes, products, and other physical phenomena as various factors are altered in the environment. Too often, data analysis, modeling, and inference are given too much emphasis at the expense of the activities that embrace the planning and execution of experiments.

It is assumed that valid and meaningful data are available either from passive observation of the process or from purposeful experiments and that statistical methods embrace the analysis of such data. It is, however, the planning or design stage leading toward the collection of data that is critical and needs to receive more attention, and it is really here that a statistical approach to the design of experiments is so important. If experiments have been designed and conducted properly, the analysis is usually straightforward and often quite simple, given the appropriate tools.

The purpose of most experimental work is to discover the direction(s) of change which may lead to improvements in both the quality and the performance of a product or process. Too often in the not too distant past, there was a tendency to conduct studies farther downstream, at the process, while the use of design of experiments was less commonly embraced by the engineering community for product design purposes.

Today, the concurrent or simultaneous engineering of products and their attendant processes is receiving widespread attention. Mathematical modeling, computer simulation, and the

associated use of designed experiments all play a central role in the design activity. In investigating the variation in performance of a given process, attention must focus on the identification of those factors which, when allowed to vary, cause performance to vary in some way. Some of these factors are qualitative in nature, sometimes referred to as categorical variables, while others are quantitative in nature with the inherent ability to change continually.

4.1 Definitions

4.1.1 Experimental Variables: The two main variables in an experiment are the **Independent and Dependent (or Response) variable**. An independent variable is the variable that is changed or controlled in a scientific experiment. It represents the cause or reason for an outcome. The **dependent or response** variable is 'dependent' on the independent variable. As the experimenter changes the independent variable, the effect on the dependent variable is observed and recorded. A change in the independent variable most likely will cause a direct change in the dependent variable. The effect on the dependent variable is measured and recorded.

4.1.2 Experiments: Not all study is research, nor is all research experimental. A true experiment is one in which certain independent variables are manipulated and their levels assigned at random in order to determine their effect on one or more response variables. The notions of ***manipulation*** and ***randomization*** are essential for a true experiment, in order to infer cause and effect.

4.1.3 Quasi-Experiments: There are experiments where randomization is not possible. For example, in an attempt to determine the effect of two methods of in-plant instruction, it may not be physically possible to randomly assign the workers to the two instruction types. Classes may have been formed by shifts or by division or operation. In this case, we can use a flip of the coin to determine which shift or group goes to which method of instruction. It is important, however, in this type of experiment, to show that the groups or classes are similar. This could be done by randomly assigning the individuals to the different groups.

4.1.4 Ex-Post-Facto Research: Ex-post-facto research is one where the experimenter has little to do with the independent variables. The independent variable would already have been acted upon and the researcher simply examines the results or the effect on the response. The researcher only studies what is available and attempts to make some inference. For example, in the study of the effect of rainfall and sunshine on crop yield on a plot of land, rainfall as an independent variable cannot be manipulated or randomized as is required in a typical experiment. As a result, inferences are dangerous in **ex-post-facto** research, except when competing hypothesis have been ruled out.

4.1.5 Correlation and Regression: In **Regression,** levels of the independent (constant) variables are set and then observations are made on the response variable. The purpose of course is to find an equation or a functional form relating y to specified ranges of the independent variables. In ***Correlation*** **studies,** the variables in pairs(x,y) are both random**.** Statistics are computed to determine the strength of the relationship between x and y. A strong correlation does not necessarily

imply a cause and effect relationship. Thus, regression is used in true experiments where the independent variables are manipulated and randomized, whereas correlation is used in ex-post-facto studies where it is desirable to find the strength of the relationship between variables.

4.1.6 Independent Variables: There are three ways to handle independent variables. They can be rigidly controlled; that is, the variables remain fixed throughout the experiment. They can be manipulated or set at levels of interest. They can also be randomized in which the order of experiments is randomized to average out the effect of the variables that cannot be controlled (noise).

4.2 Phases of an Experiment

There are three important phases of a planned experiment: planning, design, and analysis.

1. **Planning:** Clear statement of the problem and choice of measurable response/dependent variable. Choice of treatment/factors (independent variables). Are the factors quantitative (temperature) or qualitative (operator), and are they fixed or random? How will the data be collected and measured?
2. **Design** How is the data collected? How many observations are to be taken (n)? How large of a difference is to be detected? What is the method and order of randomization? What is the mathematical method to describe the experiment and what hypothesis will be tested?
3. **Analysis:** Data collection, Computation of statistics, Interpretation of results.

Important Issues in Planned Experiments

5.1 The Notion of the Mathematical Model

A fundamental task in the design of experiments is that of selection of the appropriate arrangement of test points within the real number space defined by the independent variables. Although many different considerations must come into play in selecting a test plan**,** none can be more fundamental than the notion of the mathematical model. Suppose that we are interested in a system involving a mean response, η, that depends on the input variables $x_1, x_2, \ldots, x_n$. Then one could write that: $\eta = f(x_1, x_2, \ldots, x_n: \theta_1, \theta_2, \ldots, \theta_k)$.

That is, the mean response, η, could be expressed as a mathematical function, f, with independent variables $x_1, x_2, \ldots, x_n$ and a set of parameters $\theta_1, \theta_2, \ldots, \theta_k$. Sometimes we know enough about the phenomenon under study that we can use theoretical considerations to identify the form of the function f. For example, a chemical reaction may be described by a differential equation which, when solved, produces a theoretical relationship between the dependent and independent variables. Even if the form of f is known, the values of the parameters $\theta_1, \theta_2, \ldots, \theta_k$ would generally not be. The observed data y are therefore represented as:

$$Y = f(x_1, x_2, \ldots, x_n: \theta_1, \theta_2, \ldots, \theta_k) + \varepsilon_{ij} \ ,$$

where ε_{ij} is the experimental error of the observation and is assumed to be Normal, Independent and Identically Distributed NIID(0, σ^2).

The problem is that in most situations, little is known about the underlying mechanisms of the process. In most physical situations, we can say that *f* represents a relatively smooth response function. But without knowledge of the physical mechanisms, how can we proceed to explore the response surface with the ultimate goal, perhaps, of finding the values for $x_1, x_2, \ldots, x_n$ that optimize the mean response η? When we have limited knowledge, we must rely on empirical models that serve to approximate the true, but unknown, model describing relationships through the data, for example:

$$\hat{\eta} = b_0 + b_1 x_1 + b_2 x_2$$
$$\hat{\eta} = b_0 + b_1 x_1 + b_2 x_2 + b_{12} x_1 x_2$$
$$\hat{\eta} = b_0 + b_1 x_1 + b_2 x_2 + b_{12} x_1 x_2 + b_{11} x_1^2 + b_{22} x_2^2$$

It is important to plan experiments in such a way that several specific empirical forms can be examined and the best can be determined through the data once the experiment has been completed. As an example, let us suppose that *f* can be represented, for a certain situation, by a simple and flexible graduating polynomial, say:

$$\hat{\eta} = b_0 + b_1 x_1 + b_2 x_2 + b_{12} x_1 x_2 + b_{11} x_1^2 + b_{22} x_2^2$$

If we can fit this function to data obtained from an experiment, we can use this equation to more clearly visualize and, hence, explore the complete nature of the response surface (e.g., perform optimization and sensitivity analysis.

Whether explicitly recognized as such or not**,** most experimental studies are aimed either directly or indirectly at discovering the relationship between some performance response and a set of candidate variables which influence that response. In most studies, the experimenter begins with a tentative hypothesis concerning the plausible model form(s) to be initially entertained.

We must then select an experimental design that has the ability to produce data capable of:

1. Fitting the proposed model(s) (i.e., estimate the model parameters).
2. Putting the model in jeopardy so that inadequacies in the model can be detected through analysis.

The second consideration is of particular importance in the above scenario. iWe must ensure that, through a series of iterations, the most appropriate model can be determined, even while others may be shown to be less viable through the available data. For example, if a quadratic relationship between temperature and reaction time in a chemical process is suspected, an experiment that examines the process at only two temperature levels (data points) would be inadequate to reveal this functional possibility. However, an experiment that was run at three levels of temperature would allow this functional possibility to be considered and explored. On the other hand, an experiment that has five levels of temperature would be unnecessary and inefficient if the relationship was truly

quadratic. In general, the experimental design should be responsive to the mathematical model being proposed.

The Effects of Changes in the Independent Variables

Often, the variables of importance that govern a process are not clearly known a priori. Thus, it is important to examine the effects of several variables, together and independently, to determine the true change in each of the variables. Furthermore, it is important to determine how a variable effect varies with the given states of the process (i.e., with respect to other variables). When this is the case, the design itself, namely, the arrangement of the tests, becomes a very significant factor.

Suppose, in a given situation, that four variables seem important. In a one-variable-at-a-time situation, one approach to experimentation would be to hold three of the variables constant while varying only one variable. The result of such an example could look something like figure 7.

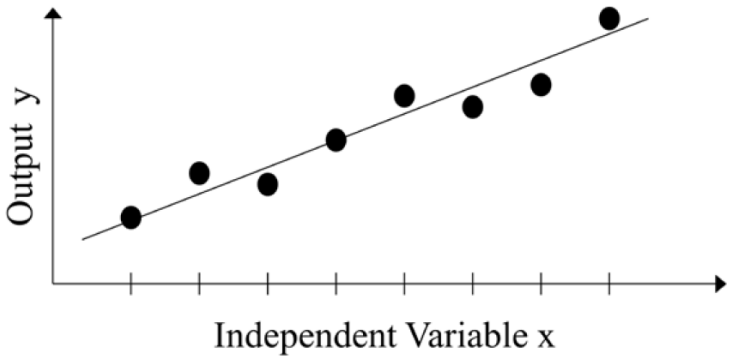


Figure 8 One-variable-at-a-time Experiments

This process would then be repeated for each of the other variables. There are several problems with this one-variable-at-a-time approach. One of the problems has to do with the fact that most variables do not influence the response of interest independent of the others. The following example will illustrate this possibility.

Suppose two factors, temperature and pressure**,** are thought to affect the reaction time of a chemical process. Two experiments, I and II, were run to understand how each factor affects reaction time. Using the one-variable-at-a-time approach**,** the results could be entirely different, due to selection of the different fixed levels for the factors that are not under consideration during the test. Figure 8 illustrates such possible outcomes as a result of the choice of different settings of these two factors between the two experimental runs. Figure 8 shows that the effect of temperature depends on the level of pressure and vice versa. This behavior connotes the idea of the interaction

of the variables. It sometimes also is a problem of multi-collinearity and/or correlation. In summary, the one-variable-at-a-time experimental approach presents problems. Specifically:

a). It generally gives rise to a large number of tests, resulting sometimes in too many levels of each factor.

b). It does not recognize variable interdependencies, which results in a highly conditional or narrow interpretation.

c). It promotes a systematic test sequence in which it is possible to have unknown effects that change during the experiment, thereby biasing the results and thus making the validity questionable.

d). It does not lend itself well to some important experimental schemes such as blocking. In any planned and well-designed experiment, the aim is to obtain an estimate of chance variation that is not inflated by assignable causes. A way to solve this problem this is to utilize the scheme of blocking in which the experiment is repeated in **blocks** where the known sources of variability are held fixed in each block, but vary from **block** to **block**.

The Effect of Noise in An Experiment

The experimental study of any phenomenon is made difficult by the presence of noise, sometimes called experimental error, which is different from natural or chance variability. Typically, each process has many more variables than are identified during each experiment. Thus, some of the factors not under study do vary during the experiment. Although the variation of any one of such factors may produce a very small change in the performance measure under study, in the greater context of the overall experiment, the fact that many such forces are at work simultaneously produces a noticeable and sometimes sizable level of system noise. When the system is stable, these are the forces of common-cause. Also, such variation,referred to as *white noise* to emphasize its random nature, may cloud or mask the effect of changes in the factors under study in an experiment. Some noise sources in the system may be of a more spurious nature. Such sporadic disturbances, sometimes of a sizable magnitude, can further contribute to clouding the results of experiments and may even bias the results, especially if their occurrence is systemic in nature. Such structured or systemic variation is sometimes referred to as *colored noise.*

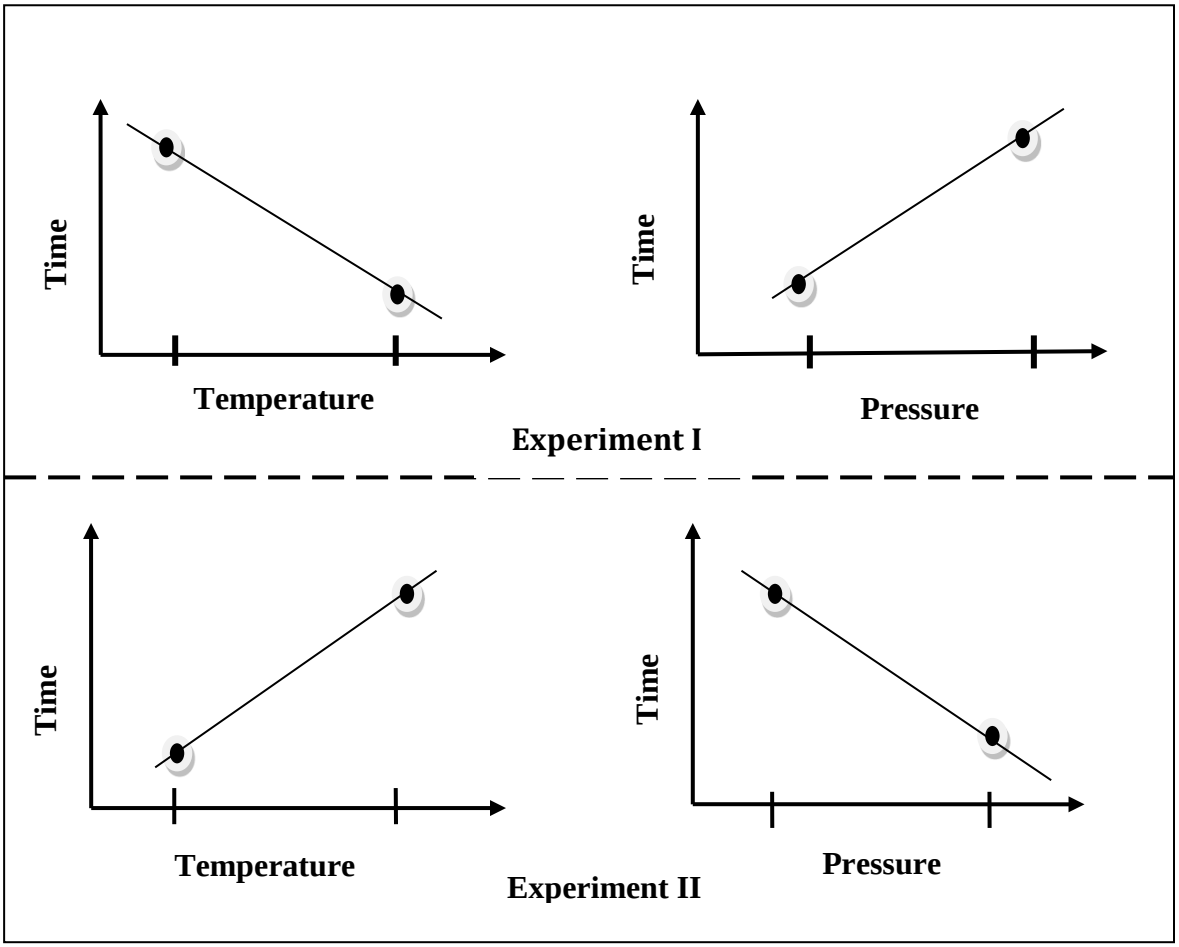


Figure 9 A Variable-at-a-Time Design

7.1 Experimental Design Strategies to Cope with Noise

Some specific design strategies that deal with the effect of noise include the following:

a). **Reduce the effect of Systemic Bias.** Counteract the effect of unknown systemic variation in the experiment by the use of randomization of the tests so that such variation is uniformly and randomly distributed across the trials conducted.

b). **Replication.** One of the most important statistics that any experimental designer would want to have is the estimate of pure error, which measures chance variation. A way to think about pure error is the following. Assume that it is possible to take readings at certain experimental points (say temperature settings, pressure settings) during an experiment. This means that if there were no other issues related with the design, the measurements obtained from the different readings taken at the same settings would be exactly the same. The only reason they would be different would be

because of chance variation or pure error. Thus, if we want an estimate of pure error, then replication is one of the approaches to data acquisition that would make that possible.

c). Identify and Isolate Known Sources of Variability through Blocking. In some production schemes, items are produced in blocks. For example, it may not be physically possible to produce all the required demand in a single shift. Thus, different production quantities might be scheduled in different shifts or days, depending on the size of the production quantity or demand. For example, a company has a certain lathe to produce a special, part but the capacity of the machine is limited so that it is not possible to produce all of the runs of the production at one run or setup. In that case, the production quantity would be broken up into blocks (days, shifts, etc) and analyzed as such.

d). Confirmatory Test. Include confirmatory testing as part of the experimental strategy. The ultimate value of designed experimentation lies with the "persistence of effects" when improvement opportunities are revealed through the experiment. Hence, it is important to run additional trials under specific conditions determined from the analysis in order to verify the improvement opportunities revealed by the experiment.

Single Factor Design of Experiments

A single factor experiment is one where the focus is only on one factor (quantitative or qualitative). All other factors are held constant. In single factor experimentation, we represent the single dependent variable and the response variable with a linear model:

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij} \quad \begin{cases} i = 1,2,...,a \\ j = 1,2,...,n \end{cases}$$

Where:

y_{ij} = $(ij)^{th}$ observations

μ = The overall mean of all the observation

τ_i = The i^{th} treatment effect

ε_{ij} is the error due to other sources of variation other than the treatment effect.

A One-Way ANOVA (Analysis of Variance) is a statistical technique for a single factor experiment with several levels (or treatments) of the single factor (usually more than three levels) in which we test to see if the different levels of the single factor (or treatments) are significantly different based on the responses obtained from the experiments. It tests whether the value of a single variable differs significantly among three or more levels of a factor. We can say we have a framework for one-way *ANOVA* when we have a single factor with three or more levels and multiple observations at each level. There seems to be a lot of emphasis on the levels of the treatment. Why? As we will find out later, the levels of a factor determine the degrees of freedom

for the effect which makes it possible to conduct a test of significance using the degrees of freedom as the divisor.

Consequently, if the number of levels of a factor is two, the degrees of freedom is *a*-1 or (2-1) which is 1. The degrees of freedom should have a value of at least two or higher, if at all possible. Later we will see the significance of this rule of thumb and how it is implemented on an ANOVA table.

On the *ANOVA* table, the Total Sum of Squares (SSTotal) is parttioned into the Sum of Squares due to treatment, the Sum of Squares due to blocks or restrictions and the Error Sum of Squares. In other words, the Total Sum of Squares is the sum of all the Sum of Squares of all the admissible terms or elements in the *ANOVA* table when properly adjusted. In certain cases the Sum of Squares certain elements cannot be adjusted and they are not considered as part of the Total Sum of Sqaures and as such they cannot be tested for significance.

As we shall see from the *ANOVA* table, the Sum of Squares Total (*SST*)**,** and the Sum of Squares for each element (treatment, block, error) are used to compute the Mean Square (*MS*). The *MS* for each element on the *ANOVA* table is obtained by dividing each Sum of Squares by the correspondigng degree of freedom. To test for significance, the *MS* of each term in the model is divided by the *MS(Error).* The ratio of an *MS* to *MS*(*Error*) is the computed F-ratio. If the F-ratio is greater than the F-value from the table, then we have significance. If it is less than one (i.e., if the Fratio <1)then that particular element or tem on the *ANOVA* table is not significant.

Total: df(Total)=*N*-1, Treatment (dfTr)= *a*-1 , Error (Residual df Error)=(*N*-1)-(*a*-1).

Table 4 Data Layout for Single Factor Design

Treatment Levels	Observations					Totals	Average
1	y_{11}	y_{12}	•	•	y_{1n}	$y_{1\bullet}$	$\bar{y}_{1\bullet}$
2	y_{21}	y_{22}	•	•	y_{2n}	$y_{2\bullet}$	$\bar{y}_{2\bullet}$
•	•	•	•	•	•	•	•
•	•	•	•	•	•	•	•
a	y_{a1}	y_{a2}	•	•	y_{an}	$y_{a\bullet}$	$\bar{y}_{a\bullet}$
						$y_{\bullet\bullet}$	$\bar{y}_{\bullet\bullet}$

Note that the treatment effect of fixed preselected levels of the independent variable is the focus here. Hence, the results cannot be extended beyond the "**a**" levels. When the number of levels of the treatment is fixed then we have a *Fixed Effects Model*. If on the other hand the treatment levels are chosen as a random sample from a larger population, then the treatment effect τ_i is now a random variable. Here the knowledge about particular τ_i's that are been investigated is useless. Instead, we

test hypothesis about the variability about τ_i and try to estimate this variability. This is called *Random Effects Model*.

8.1 Analysis of the Fixed Effects Model

Let τ_i = deviation from the overall mean for i^{th} treatment. Hence, $\sum_{i=1}^{a}\tau_i = 0$

$$y_{i.} = \sum_{j=1}^{n} y_{ij}, \quad \overline{y_{i.}} = \frac{y_{i.}}{n}$$

$$y_{..} = \sum_{i=1}^{a}\sum_{j=1}^{n} y_{ij}, \quad \overline{y_{..}} = \frac{y_{..}}{a \times n} = \frac{y_{..}}{N}$$

Hypothesis:

For at least one pair (i,j)

$$H_0 : \mu_1 = \mu_2 = ... = \mu_a = \mu., \; H_1 : \mu_i \neq \mu_j$$

$$H_0 : \tau_1 = \tau_2 = ... = \tau_a = 0, H_1 : \tau_i \neq 0$$

The equivalent Hypothesis is the Decomposition of the Total Sum of Squares:

$$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{n}\left(y_{ij} - \overline{y_{..}}\right)^2 = n\sum_{i=1}^{a}(\overline{y_{i.}} - \overline{y_{..}})^2 + \sum_{i=1}^{a}\sum_{j=1}^{n}\left(y_{ij} - \overline{y_{i.}}\right)^2$$

$$= SS_{Treatments} + SS_E$$

$$F_0 = \frac{MS_{treatment}}{MS_E} \sim F_{a-1, N-a}$$

Reject if:

$$H_0 : \tau_1 = \tau_2 = ... = \tau_a = 0, \quad H_1 : \tau_i \neq 0$$

$$F_0 > F_{\alpha, a-1, N-a}$$

Table 5 One-Way ANOVA

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0
Between Treatments	$SS_{treatments}$	a-1	$MS_{treatments}$	$F_0 = \frac{MS_{treatment}}{MS_E}$
Error (within treatments)	SS_E	N-a	MS_E	
Total	SS_T	N-1		

Example: A multinational company makes cutting fluids that are designed for use with micro-lubricant applicators and are used in applications such as drilling, tapping and sawing. These non-misting applicators dispense individual droplets of fluid directly to the cutting tool. A customer is in the process of making a decision regarding which cutting fluid will best suit its production operation. Five different types of fluids are available. Suitability is measured by temperature rise and tool life. In this example, we are measuring tool life in hours. Five oil types (A, B, C, D, E) are available. The engineer has decided on a sample size of 8 for each oil type. Hence, this is a single-factor experiment with five levels (a=5), and eight replicates(n=8).

The appropriate mode for this one-way design is a completely randomized version of the one-way or single factor *ANOVA.* Before we actually run the experiment, we must determine the order or the sequence. Since this is a completely randomized design, we will assign the sequence of tests or experimentation randomly. To do this, we first note that there will be a total of forty (5x8) data points or test results, that is *N=40.* We will label each potential observation from 1 to 40, starting with the first oil, type A, through the last oil, type E. This labeling does not accord any importance or significance to any oil type because we could easily have started from oil type E to oil type A. Table 6 shows the oil type and the experimental run number. Please note that the ANOVA is about the significance test based on the F-statistic. The test involves the error means square and the Treatment means square, both are measures of variance.

Table 6 Oil Type and the Experimental Run Number

Oil Type	Experimental Run No.							
A	1	2	3	4	5	6	7	8
B	9	10	11	12	13	14	15	16
C	17	18	19	20	21	22	23	24
D	25	26	27	28	29	30	31	32
E	33	34	35	36	37	38	39	40

Using the random number generator (RAND()) in EXCEL, we then generate random numbers between 1 and 40 inclusive to determine the test sequence(RAND ( ) * (b−c) + c)
This function will return random numbers from the interval [c,b] - greater than or equal to c, and less than b. Based on this, for example, if the first random number generated is 5, then first experimental run will be carried out using oil type A. If the second random number generated is 22, then the second run will use oil type C. Table 7 shows the oil type, the corresponding run number and the test sequence, as determined by the random number

Table 7 Oil Type, Run Number and Test Sequence

S/N	A	B	C	D	E
1	4.435	3.635	3.845	3.611	4.722
2	3.341	3.683	3.378	3.796	4.102
3	3.918	3.587	5.274	3.471	5.013
4	5.031	3.852	3.333	3.265	3.682
5	3.843	5.091	3.646	4.479	4.156
6	4.183	4.287	4.528	4.662	3.652
7	3.236	4.099	4.381	2.999	4.195
8	4.214	3.911	3.761	5.044	3.865
Totals	32.201	32.145	32.146	31.327	33.387
Mean	4.025	4.018	4.018	3.916	4.173
Std Dev	0.583	0.495	0.664	0.729	0.481

$\hat{\mu}_A = 4.025$
$\hat{\mu}_B = 4.018$
$\hat{\mu}_C = 4.018$
$\hat{\mu}_D = 3.916$
$\hat{\mu}_E = 4.173$
$MS_F = 0.357386$

Table 8 Test Results for the Five Oil Types

	Run Number	Oil Type	Test Sequence	Run Number	Oil Type	Test Sequence	Run Number	Oil Type
1	5	A	15	20	C	29	9	B
2	22	C	16	24	C	30	31	D
3	36	E	17	29	D	31	4	A
4	1	A	8	40	E	32	18	C
5	10	B	19	37	E	33	25	D
6	30	D	20	35	E	34	32	D
7	2	A	21	28	D	35	27	D
8	14	B	22	17	C	36	26	D
9	19	C	23	21	C	37	15	B
10	23	C	24	12	B	38	3	A
11	11	B	25	16	B	39	7	A
12	39	E	26	38	E	40	8	A
13	33	E	27	6	A			
14	34	E	28	13	B			

Table 8 is the actual experimental results for each oil type together with their means and standard deviation. Based on these results we can perform an ANOVA (table 9) to determine the significance

of the of the treatment effect. **Any time the statistic of the significant test, namely the value of the F-statistic is less than 1, then the test under consideration is not significant**.

Table 9 ANOVA for the Single Factor Model

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0
Between Treatments	0.271069	4	0.06776	<1 (=0.1896)
Error (within treatments)	12.5085	35	0.35738	
Total	12.77957	39		

Since F_0 is less than 1, then we cannot reject H_0.

$$H_0 : \mu_1 = \mu_2 = ... = \mu_a = \mu$$

This implies that the different oil types are not significantly different, therefore, everything else being equal, we should select the one with the lowest processing time, oil type D.

8.1.1 Estimation of Model Parameters and Confidence interval

The single factor model: $y_{ij} = \mu + \tau_i + \varepsilon_{ij}$

An estimate of the overall mean and the treatment effect can be given as:

$$\hat{\mu} = \overline{y..}\ , \quad Let: \hat{\tau}_i = \overline{y_i}. - \overline{y..}$$

Then, an estimate of μ_i would be: $\mu_i = \mu + \tau_i \qquad i = 1,2,...,a$

If we assume that errors are normally distributed, then we can say that the mean response is normal, independent and identically distributed where: $\overline{y_{i.}} \sim NID(\mu_i, \sigma^2/n)$

Thus, if σ^2 is known, then the Confidence Interval (CI) can be established using the normal distribution. We can use MS_E as an estimator of σ^2, if we assume that the errors are normally distributed. In this case we will base the CI on the student- t distribution

$$\left[\overline{y_{i.}} \pm t_{\alpha/2, N-a}\sqrt{\frac{MS_E}{n}}\right]$$

For the following problem, we can establish a CI for mean value of oil type D. $\overline{Y}_D = 3.916$ for a 90% confidence interval t (0.05, 35) = 1.645, $S_D = \sqrt{\frac{0.3574}{8}} = 0.2114$

$$3.916 \pm 1.645\ (0.2114) = [3.916 \pm 0.3477] = [4.2637,\ 3.5683]$$

$$P[\mathrm{UCL}_D \le \mu_D \le LCL_D] = 90\%$$

$$P[3.5683 \le \mu_D \le 4.2637] = 90\%$$

We may also be interested in looking at the difference between two pairs of means. Assume that we want to look at the difference between the lowest and highest response, D and E. The 100(1-α)% C.I. on the difference of any two treatments means, μ_E and μ_D would be:

$$\overline{y_{i.}} - \overline{y_{j.}} \pm t_{\alpha/2, N-a}\sqrt{\frac{2MS_E}{n}}$$

Why is the variance of the difference between the two means equal to $2MS_E/n$? Recall that for a random variable X and constant c, Var $(cX) = c^2$Var(X)
Hence Var$(-X)$ = $(-1)^2$Var(X) = Var(X)

$$Var(\bar{y}_{i.} - \bar{y}_{j.}) = Var(\bar{y}_{i.}) + Var(\bar{y}_{j.})$$
$$Var(\bar{y}_{i.}) = \frac{MS_E}{n},\ Var(\bar{y}_{j.}) = \frac{MS_E}{n}$$
$$Var(\bar{y}_{i.} - \bar{y}_{j.}) = \frac{MS_E}{n} + \frac{MS_E}{n} = \frac{2MS_E}{n}$$
$$S_{\bar{y}_{i.} - \bar{y}_{j.}} = \sqrt{\frac{2MS_E}{N}}$$
$$P(-0.2347 \le \mu_{\bar{y}_{i.} - \bar{y}_{j.}} \le 0.7487) = 90\%$$

$$Var(cX) = c^2 Var(X)$$
$$Var(-X) = (-1)^2 Var(X)$$
$$V(a-b) = Var(a) + (-1)^2 Var(b)$$

$$\overline{y_{i.}} - \overline{y_{j.}} \pm t_{\alpha/2, N-a}\sqrt{\frac{2MS_E}{n}} = (4.173 - 3.916) \pm 1.645\sqrt{\frac{2(0.3574)}{8}} = [0.257 \pm 0.4917]$$

Note that since the confidence interval included zero, the difference between the two means is not significant. We can demonstrate this by looking at the test of hypothesis between the two means:

$$H_0 : \mu_E - \mu_D = 0,\ H_1 : \mu_E - \mu_D \neq 0,$$

Let α=0.1, α/2=0.05, MS_E=0.3574, $t_{0.05, 32}$=1.645

Test Statistic:
$$t = \frac{(\mu_E - \mu_D) - 0}{\sqrt{\frac{2MS_E}{n}}} = \frac{(0.257)}{0.2989} = 0.8598 < 1$$

Reject if: $|t| > t_{\alpha/2, 32}$, therefore reject since **0.8598 <1.645**.

As noted earlier, if the numerical value of an F-test is less than unity, then that test statistic is not significant. It is instructive to also note that the result from a Confidence Interval analysis will always coincide or align with that from the test of hypothesis, with regard to the resulting decision. Recall that the two approaches are simply different pathways to perform statistical inference so it should not be surprising that they lead to the same decision.

8.1.2 Model Adequacy Checking

A major assumption we made in formulating or proposing the model is that it represents the underlying relationship in the data or the data structure. That means that the observations are adequately described by the model. If the model is not adequate or appropriate, then the proposed model is of no use and the test for the equality of means is of little use. A way to test for the

adequacy of the model is to look at the residuals. If the model adequately fits the data, then there should be no discerning structure in the residual data.. The proposed linear model is given as:

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij}\ , \qquad where\ \varepsilon_{ij}\ is\ NID(0, \sigma^2)$$

Examination of residuals can be used to investigate violations of the above assumption as well as model accuracy. The residual for observation *j* in treatment *i* is: $\varepsilon_{ij}, where\ \varepsilon_{ij} = y_{ij} - \hat{y}_{ij}$

Where, $\hat{y}_{ij}$ is an estimate of y_{ij} and is given by: $\hat{y}_{ij} = \hat{\mu} + \hat{\tau}_i = \bar{y}_{..} + (\bar{y}_{i.} - \bar{y}_{..}) = \bar{y}_{i.}$

$$Thus: \varepsilon_{ij} = y_{ij} - \hat{y}_{ij} = y_{ij} - \bar{y}_{i.}$$

8.1.3 Plot of Residuals

A scatter plot of the residual (table 10 and figure 9) reveals no discernible structure. This means that our assumption of a linear model and the error being normally, identically, and independently distributed (NIID) is sustained. That means that the errors can be assumed to be white noise.

8.1.4 The Normality Assumption.

A useful procedure for checking the normality assumption is to construct a normal probability plot of residuals. First, calculate the residuals. The residuals are ranked in ascending order and the probability associated with each rank is plotted. We use thefollowing equation to compute the probability of occurrence at the i^{th} rank:

$$P_i = \frac{(i-0.5)}{N} \quad OR \quad P_i = \frac{(i-0.3)}{N+0.4}.$$

The plot will be probability, P_i versus the ranked residuals (e_{ij}). If the plot resembles a straight line, the assumption of normality is not violated. Points falling far above the line are called outliers. The presence of one or more outliers can seriously distort the ANOVA. Potential outliers call for serious investigation. If the errors are truly normal and identically distributed, then the normal plot will be a straight line (see figure 10).

Table 10: Residuals

A	B	C	D	E
-0.415	0.383	0.173	0.305	-0.549
0.679	0.335	0.640	0.120	0.071
0.102	0.431	-1.256	0.445	-0.840
-1.011	0.166	0.685	0.651	0.491
0.177	-1.073	0.372	-0.563	0.017
-0.163	-0.269	-0.510	-0.746	0.521
0.784	-0.081	-0.363	0.917	-0.022
-0.194	0.107	0.257	-1.128	0.308
$\bar{y}_A - y_{ij}$	$\bar{y}_B - y_{ij}$	$\bar{y}_C - y_{ij}$	$\bar{y}_D - y_{ij}$	$\bar{y}_E - y_{ij}$

Figure 10 Residual Scatter Plot

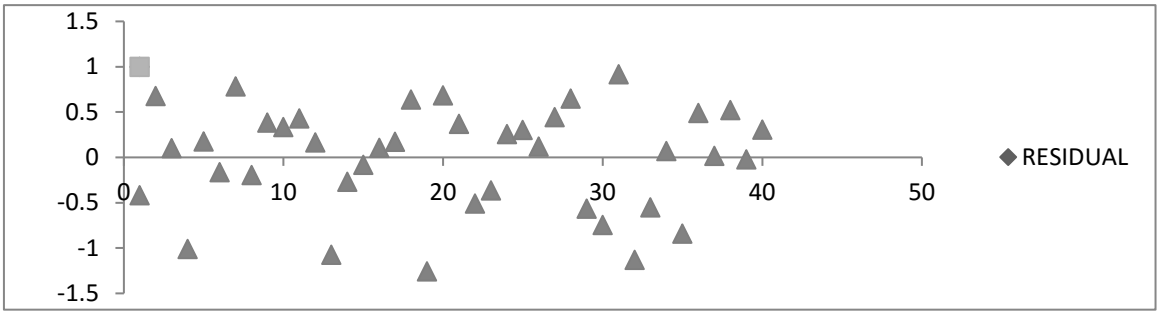


Table 11 Ranked Residuals

s/n	e_{ij}	P_i	s/n	e_{ij}	P_i
1	-1.256	0.0125	21	0.12	0.5125
2	-1.128	0.0375	22	0.166	0.5375
3	-1.073	0.0625	23	0.173	0.5625
4	-1.011	0.0875	24	0.177	0.5875
5	-0.84	0.1125	25	0.257	0.6125
6	-0.746	0.1375	26	0.305	0.6375
7	-0.563	0.1625	27	0.308	0.6625
8	-0.549	0.1875	28	0.335	0.6875
9	-0.51	0.2125	29	0.372	0.7125
10	-0.415	0.2375	30	0.383	0.7375
11	-0.363	0.2625	31	0.431	0.7625
12	-0.269	0.2875	32	0.445	0.7875
13	-0.194	0.3125	33	0.491	0.8125
14	-0.163	0.3375	34	0.521	0.8375
15	-0.081	0.3625	35	0.64	0.8625
16	-0.022	0.3875	36	0.651	0.8875
17	0.017	0.4125	37	0.679	0.9125
18	0.071	0.4375	38	0.685	0.9375
19	0.102	0.4625	39	0.784	0.9625
20	0.107	0.4875	40	0.917	0.9875

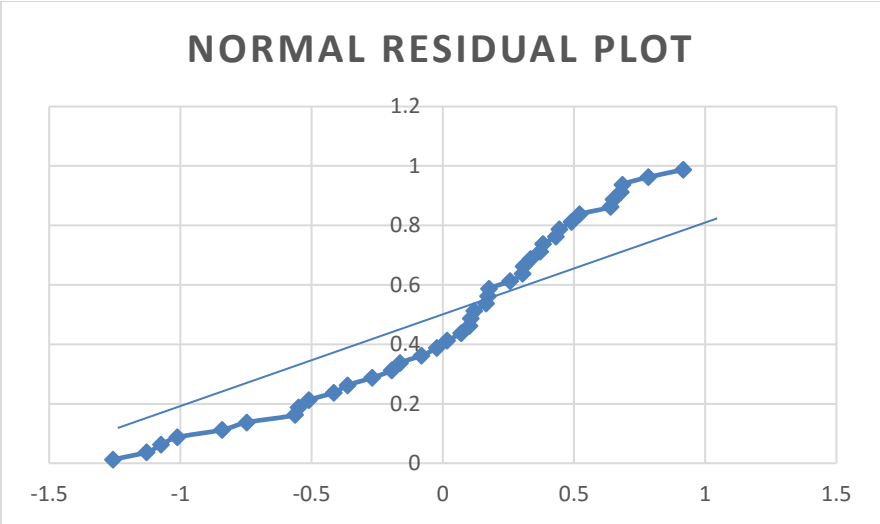


Figure 2 Normal Plot of Residuals

8.2 Comparison among Treatment Means

After H_0 is rejected, we need to find exactly which means differ. For this, comparisons of groups of treatment means is quite be useful.

8.2.1 Multiple Comparison Methods: Contrasts

Some sample multiple comparison hypotheses are:

$$H_0 : \mu_4 = \mu_5, H_1 : \mu_4 \neq \mu_5 \Rightarrow y_{4.} - y_{5.} = 0$$

$$H_0 : \mu_1 + \mu_3 = \mu_4 + \mu_5,\ H_1 : \mu_1 + \mu_3 \neq \mu_4 + \mu_5 \Rightarrow y_{1.} + y_{3.} - y_{4.} - y_{5.} = 0$$

The above hypothesis could be used to investigate an appropriate linear combination of treatment totals. In general, multiple comparisons will imply a linear combination of treatment totals such as:

as: $C = \sum_{i=1}^{a} c_i y_{i.}$ with the restriction that: $\sum_{i=1}^{a} c_i = 0$

C is called a contrast. *Sum of Squares* of a contrast C is calculated as: $SS_C = \dfrac{\left(\sum_{i=1}^{a} c_i y_{i.}\right)^2}{n\sum_{i=1}^{a} c_i^2}$

which has a single degree of freedom

8.2.2 Orthogonal Contrasts

Contrasts in general or useful for multiple comparisons. Orthogonal contrasts are set prior to the experiment and are for preplanned comparisons. The contrasts are specified before actually running or acquiring the data because certain levels or combination of levels of a factor are of special interest. If on the other hand the comparison are decided upon after the data has already being collected, then such a situation is called snooping or "data snooping".

For balanced design (all sample sizes are equal), two contrasts with coefficients $\{c_i\}$ and $\{d_i\}$ are orthogonal if: $\sum_{i=1}^{a} c_i d_i = 0$

For unbalanced design (sample sizes are not equal) if $\sum_{i=1}^{a} n_i c_i d_i = 0$

For '***a***' treatments, there are maximum of ***a-1*** orthogonal contrasts and they are NOT unique. There are many ways to choose orthogonal contrasts. For example, for three treatments ***(a=3),*** there are a maximum of two (2) orthogonal contrasts.

	$y_{.1}$	$y_{.2}$	$y_{.3}$
C_1	-2	1	1
C_2	0	-1	1

$$\sum_{i=1}^{a} c_i = -2+1+1$$
$$\sum_{i-1}^{a} c_i = 0-1+1$$

To show that these contrasts are orthogonal, we multiply the coefficient of each contrast and sum. The sum should be zero. That is: (0)(-2)+(-1)(1)+(1)(1)=0.

Please note that since the orthogonal contrasts are formed from the linear combinations of the treatment means or treatment totals, the **sum total of the Sum of Squares of the Orthogonal Contrasts must equal the Sum of Squares of the treatment that is:**

$$\sum SS(Orthogonal\ \ Contrasts) = SS(Treatment)$$

In our oil type example with five levels, the maximum number of orthogonal contrast is (a - 1 =4), since a = 5 in this case). One such possible set of four orthogonal contrast is as shown in table 12.

Table 12 Four Orthogonal Contrasts for Five Treatment Levels

		Contrasts				
Hypothesis	-		$y_{.A}$	$y_{.B}$	$y_{.C}$	$y_{.D}$	$y_{.E}$
H_0: $4\mu_A=\mu_B+\mu_C+\mu_D+\mu_E$		C_1	4	-1	-1	-1	-1
H_0: $\mu_C=\mu_D$		C_2	0	0	1	-1	0
H_0: $\mu_B=\mu_E$		C_3	0	1	0	0	-1
H_0: $\mu_C+\mu_D=\mu_B+\mu_E$		C_4	0	-1	1	1	-1

Table 13 ANOVA Table for the Treatment and Orthogonal Contrasts

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0	Sig
Between Treatments	0.271069	4	0.06776725	0.189619	n.s
Orthogonal Contrasts					
$4\mu_A=\mu_B+\mu_C+\mu_D+\mu_E$	0.0003	1	0.0003	0.0008	n.s
$\mu_C=\mu_D$	0.04192	1	0.04192	0.1172	n.s
$\mu_B=\mu_E$	0.0964	1	0.0964	0.2697	n.s
$\mu_C+\mu_D=\mu_B+\mu_E$	0.1324	1	0.1324	0.3705	n.s
Error (within treatments)	12.5085	35	0.357385714		
Total	12.77957	39			

Computation of SS for the Contrasts:

C_1=4(32.201)-1(32.145)-1(32.146)-1(31.327)-1(33.387) = -0.201, $SS_{C_1} = \frac{(-0.201)^2}{8(20)} = 0.0003$

C_2=1(32.146)-1(31.327) = 0.819, $SS_{C_2} = \frac{(0.819)^2}{8(2)} = 0.04192$

C_3=1(32.145)-1(33.387) = -1.242, $SS_{C_3} = \frac{(-1.242)^2}{8(2)} = 0.0964$

C_4=1(32.146) + 1(31.327)-1(32.145)-1(33.387)= -2.059, $SS_{C_4} = \frac{(-2.059)^2}{8(4)} = 0.1324$

From the contrast analysis (see table 13) we come to the same conclusions we arrived at earlier. There are no significant variations between the different levels of factors. In other words, the oil types are not considerably different in terms of performance. Please notice that we arrived at this conclusion without even specifying the alpha level or the Type I error. Why? If the F values are less than unity anytime we carry out an ANOVA test, , we know that, regardless of the Type I error specified, the test is not significant. Please note that the combined sum of squares for the orthogonal contrasts must equal the treatment sum of squares.

8.2.3 Schefee's Test for any and all Contrast

Sometimes we do not know which mean comparisons would be needed before data has been collected from an experiment. It is usually after looking at the data, after the fact, ('data snooping') that it become clear which pair of means or linear combination of means need to be examined. Such contrasts could be as little as one or more than the number of treatment levels and thus may not be orthogonal. We know that in the case of orthogonal contrasts the maximum is (*a*-1) for ('*a*') treatment levels. When the desire is to examine any and all contrasts that are not necessarily orthogonal, the way to do that is by using an approach proposed by Scheffe's in 1953 which has come to be known as Scheffe's Test. Thus, Scheffe's method applies to the set of estimates of all possible contrasts among the factor level means; and, that number could be theoretically (but not possibly) infinite. The Scheffe's test is conducted on each contrast. The estimated variance or (the Mean Square Error from the ANOVA table) for the contrast is given by:

$$S_{C_k}^2 = MS_E \sum_{i=1}^{a} \frac{C_{ik}^2}{n_i} \Rightarrow S_{C_k} = \sqrt{MS_E \sum_{i=1}^{a} \frac{C_{ik}^2}{n_i}}$$

For the Scheffee's test, the critical value to be compared to the contrast *C* is given by:

$$S_{\alpha,k} = S_{C_k} \sqrt{(a-1)F_{\alpha,(a-1),(N-a)}}$$

Example for two contrasts(*k*=2):

H_0: $4\mu_A = \mu_B + \mu_C + \mu_D + \mu_E$, Hence, $C_1 = 4\bar{y}_{A\bullet} - \bar{y}_{B\bullet} - \bar{y}_{C\bullet} - \bar{y}_{D\bullet} - \bar{y}_{E\bullet} = -0.201$

H_0: $\mu_C + \mu_D = \mu_B + \mu_E$, Hence, $C_2 = \bar{y}_{C\bullet} + \bar{y}_{D\bullet} - \bar{y}_{B\bullet} - \bar{y}_{E\bullet} = -2.059$

Reject if $|C_K| > S_{\alpha,k}$, α=0.05, $n_A = n_B = n_C = n_D = n_E = 8$

$$S_{C_k} = \sqrt{MS_E \sum_{i=1}^{a} \frac{C_{ik}^2}{n_i}} = \sqrt{(0.3574)\left(\frac{20}{8}\right)} = \sqrt{(0.8935)} = 0.9452$$

$$S_{C_k} = \sqrt{MS_E \sum_{i=1}^{a} \frac{C_{ik}^2}{n_i}} = \sqrt{(0.3574)\left(\frac{4}{8}\right)} = \sqrt{(0.1787)} = 0.4227$$

For C_1, $|C_1| = 0.201 < 0.9452\left(S_{C_1}\right) \Rightarrow$ Do not Reject

For C_2, $|C_2| = 2.059 > 0.4227\left(S_{C_1}\right) \Rightarrow$ Reject. Those pairs of means are not equal

Single Factor: Restrictions on Randomization (Randomized Block Designs)

Most experiments contain nuisance factors that produce variability, ultimately affecting the response variable by inflating the MS_E(the Error Variance). This is of no particular interest to the engineer. If a nuisance factor is known but uncontrollable, then an approach that is beyond the scope of this course, namely ANCOVA--Analysis of Covariance, may be used to extricate the effect of the nuisance factor from the results. If a nuisance factor is unknown and uncontrollable, then randomization can be used to mitigate or even out its effect on the results. If, however, the nuisance factor is known and controllable or can be manipulated, then the scheme of "Blocking" can be used to reduce or eliminate its effect on the response or the estimate of the error variance.

In the single factor, completely randomized design we looked at earlier, the experimental units were randomly assigned. In such a case, any variation from sample to sample (within sample variation) appeared in the error variation as measured by MS_E. In many cases, the variability can come from the large heterogeneity in the experimental units that results in reduced sensitivity or ability to detect treatment difference because variations across the experiment units have inflated MS_E.

As an example, due to the time required to complete an experimental trial, it would be impossible to complete the data extraction during the day shift. Thus, management decides to collect data across two shifts (day and night shifts). If the data collection is completely randomized and analyzed as such t any differences due to the different shifts will be masked because it would be lumped with the MS_E thereby inflating it. A better and more practical approach is to 'block off' the variation due to these units (shifts) by randomly varying them over a wider range in smaller or more homogeneous block (shifts) in such a way that their variability can be eliminated from MS_E. In a randomized block design **(see figure 12),** the experimenter divides treatments subgroups--blocks, such that the variability within blocks is less than the variability between blocks. This type of design that would extract the blocking effect from the true experimental error is the randomized complete block design (RCBD) in which each block 'completely' contains all the treatments. Within each block, the treatments are assigned at random. Therefore, rather than complete randomization, the blocks are used as a form of restriction on randomization. Again, a design that utilizes this strategy

of blocking is called the Randomized Block Design. There are two type, (a) Balanced Complete and (b) Balanced Incomplete design.

9.1 Randomized Complete Block Design (Balanced Complete Design)

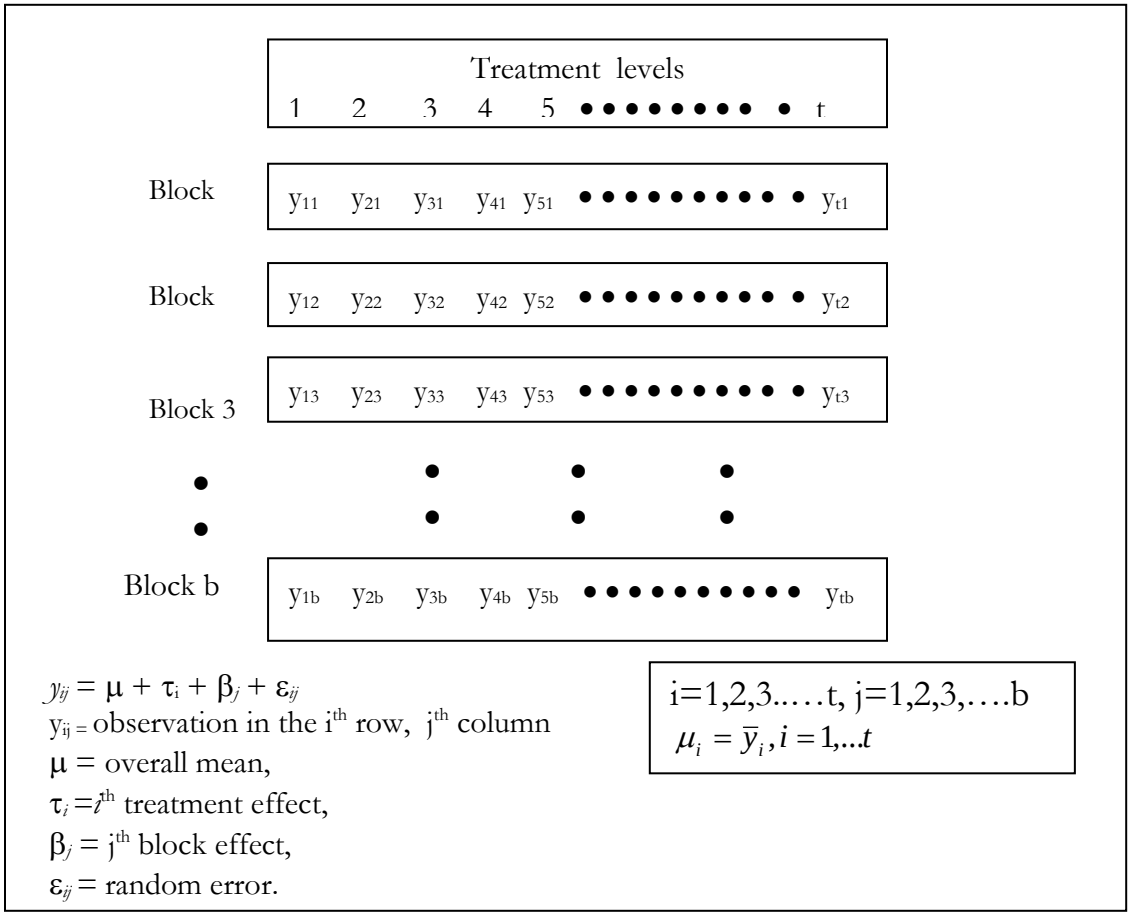
As indicated earlier, this is a design in which each level of treatment (the single factor) occurs once and only once in each block. We are interested in testing the equality of the treatment means. Thus the hypotheses of interest are: H_o: $\mu_1 = \mu_2 = \ldots = \mu_t$, H_1: at least one $\mu_1 \neq \mu_2$

To test the equality of treatment means, we would use the test statistic: $F_0 = \frac{MS_{Treatment}}{MS_E}$

Where F_0 is distributed as $F_{\alpha,(a-1),(a-1)(b-1)}$

if the null hypothesis is true, that is, if H_0 is true The critical region is the upper tail of the F distribution, and we would reject H_0 if: $F_0 > F_{\alpha,(a-1),(a-1)(b-1)}$. The analysis is better captured in the *ANOVA* table 14 as shown.

Figure 12 Layout of Complete Block Design with Randomization



Where F_0 is distributed as: $F_{\alpha,(a-1),(a-1)(b-1)}$

if the null hypothesis is true, that is, if H_0 is true, the critical region is the upper tail of the F distribution, and we would reject H_0 if: $F_0 > F_{\alpha,(a-1),(a-1)(b-1)}$.

The analysis is better captured in the (ANOVA) table as shown.

Table 14 ANOVA Table for Randomized Complete Block Design

Sources of Variation	**Sum of Squares**	**Degrees of Freedom**	**Mean Square**	**F_0**
Treatment	$SS_{\text{Treatment}} = \frac{1}{b}\sum_{i=1}^{t} y_{i.}^2 - \frac{y_{..}^2}{tb}$	t-1	$\frac{SS_{Treatment}}{(t-1)}$	$\frac{MS_{Treatment}}{MS_E}$
Blocks	$SS_{\text{Block}} = \frac{1}{t}\sum_{j=1}^{b} y_{.j}^2 - \frac{y_{..}^2}{tb}$	b-1	$\frac{SS_{Block}}{(b-1)}$	$\frac{MS_{Block}}{MS_E}$
Error (Residual)	SS_E (obtained by subtraction)	$(t-1)(b-1)$	$\frac{SS_E}{(b-1)(t-1)} = MS_E$	
Total	$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{b} y_{ij}^2 - \frac{y_{..}^2}{tb}$	$(tb-1)$ or N-1		

Example: An experiment to determine the amount of warping (mm) of copper plates was conducted in 4 different laboratories (Lab 1, Lab2, Lab3, Lab4) using four copper specimens with different percent of copper compositions (A, B,C, D).

Table 15 Data for the Randomized Complete Block Design

	SPECIMEN(Treatment)				
LABORATORY	A	B	C	D	**TOTALS**
Lab1	264	208	220	217	909
Lab2	260	231	263	226	980
Lab3	258	216	219	215	908
Lab4	241	185	225	224	875
TOTALS	1023	840	927	882	3672

$$CF = \left(Y_{..}^{\ 2}\right)/N = 842724$$

$$SS_{Total} = \sum_{i=1}^{a}\sum_{j=1}^{n} y_{ij}^2 - \frac{y_{..}^2}{N} = 7444,\ SS_{treat} = SS_{col} = \frac{1}{n}\sum_{i=1}^{a} y_{i.}^2 - \frac{y_{..}^2}{N} = 4621.5$$

$$SS_{Block} = \frac{1}{n}\sum_{j=1}^{b} y_{j.}^2 - \frac{y_{..}^2}{N} = 1468.5,\ SS_E = SS_T - SS_{treat} - SS_{Block} = 1354$$

Microsoft EXCEL was used to compute the ANOVA table. To use EXCEL for ANOVA, you will have to install 'The Analysis ToolPack'. Once you install it as an 'add-in', you pull down the Data window menu, which then takes you to Data Analysis. For this analysis, you will choose the option: 'ANOVA: Two-factors without replication.' If you want the labels to show, you will check the 'labels' button. **In this analysis the treatments are significant but the blocks are not at α=0.05 based on the computed F values on the ANOVA table. The critical value from the F-table, F(0.05; 3,9) = 3.8625.**

Table 16 Single Factor Two-way ANOVA with Restriction (block) on Randomization Below

ANOVA: One-Factor With Blocking

SUMMARY	*Count*	*Sum*	*Average*	*Variance*
Lab1	4	909	227.25	626.25
Lab2	4	980	245	368.6667
Lab3	4	908	227	430
Lab4	4	875	218.75	566.9167
A	4	1023	255.75	102.9167
B	4	840	210	368.6667
C	4	927	231.75	440.9167
D	4	882	220.5	28.33333

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Rows (Labs-Block)	1468.5	3	489.5	3.253693	0.073833	3.862548
Columns (specimen)	4621.5	3	1540.5	10.23966	0.002929	3.862548
Error	1354	9	150.4444			
Total	7444	15				

There are other Single Factor designs but with two, three or even four restrictions on randomization. The idea is to slowly and carefully get the sum of squares for the factor and the restrictions (which are really like factors in terms of computation). The ANOVA table is populated the same way and looks exactly the same except that there more terms in the table.

9.2 One Factor With More Than One Restrictions On Randomization

To wrap up this section on restrictions on randomization, we will quickly look at a situation where we have one factor with two restrictions on randomization, namely the Latin Square Design.

9.2.1 Latin Square Design. The Latin Square is a 4x4 symmetric design with **two restrictions** on randomization where each treatment occurs once, and only once, in each row and in each

column. Thus, the result is a 3-way ANOVA table (Two terms for the two blocks and one term for the Treatment)

Table 17 A 4x4 Latin Square Design

	Machines (Block)			
Position	1	2	3	4
1	C	D	A	B
2	B	C	D	A
3	A	B	C	D
4	D	A	B	C

9.2.2 The Greco Latin Square Design.

The Greco Latin Square Design is a Latin Square Design with one additional restriction on randomization. In other words, instead of two restrictions, it has three restrictions. The ANOVA scheme is the same except for one additional term in the 4-way ANOVA table. In what follows, we show the restrictions (3 at 5 levels each), the data table, the partial sums, and the Sum of Squares (SS) for each term in the model, including the SS Total. The degrees of freedom are determined the same way as always and are as shown. What is now left is to show the F values for each term in the ANOVA table to determine significance (table 18).

$$MODEL: Y_{ijkl} = \mu + T_i + T_j + T_k + T_l + \varepsilon_{ijkl}, \quad T_{\bullet\bullet\bullet\bullet} = 62, \quad CF = \frac{T^2_{\bullet\bullet\bullet\bullet\ldots}}{25} = 153.74$$

$$SS_{Treatment} = \frac{\sum T^2_{\bullet\bullet\bullet k}}{5} - CF = \left(\frac{14^2 + 12^2 + 11^2 + 12^2 + 13^2}{5}\right) - \frac{(62)^2}{25} = 1.06,$$

$$SS_{Batch} = \frac{\sum T^2_{i\bullet\bullet\bullet}}{5} - CF = \left(\frac{11^2 + 12^2 + 12^2 + 16^2 + 11^2}{5}\right) - \frac{(62)^2}{25} = 3.44$$

$$SS_{Order} = \frac{\sum T^2_{\bullet j\bullet\bullet}}{5} - CF = \left(\frac{12^2 + 10^2 + 14^2 + 13^2 + 13^2}{5}\right) - \frac{(62)^2}{25} = 1.86$$

$$SS_{operator} = \frac{\sum T^2_{\bullet\bullet\bullet l}}{5} - CF = \left(\frac{17^2 + 9^2 + 11^2 + 13^2 + 12^2}{5}\right) - \frac{(62)^2}{25} = 7.06$$

$$SS_{Total} = \sum\sum\sum\sum Y_{ijkl} - CF = (34 + 26 + 40 + 35 + 35) - 153.74 = 16.24$$

$$SS_{Error} = SS_{Total} - SS_{Treatment} - SS_{Batch} - SS_{Order} - SS_{Operator} = 2.84$$

$$DF_{Error} = (N-1) - (4+4+4+4) = 24 - 16 = 8, \; MS_E = \frac{2.84}{8} = 0.355$$

9.3 Randomized Incomplete Block Design

In some randomized block design, it may not be possible to apply the treatment in every block. An incomplete block design is one in which there are more treatments than can be put in a

single block. In a balanced incomplete block design, every pair of treatments occur the same number of times throughout the experiment. Consider a situation where we have four operators (treatment) and four working days for an assembly task with time in seconds. However, at any given day out of the four working days, the schedule calls for only three operators to be at work. In such a case, the block (day) would be incomplete because it only has three treatments in it. The data in the table has been normalized bysubtracting 850 from the actual data, just for ease of hand calculation

. **Table 18 Data layout and Computation (below) for the Greco Latin Square Design**

3 RESTRICTIONS ON RANDOMIZATION						**DF**
TREATMENT (*k*)	A	B	C	D	E	4
BATCHES (*i*)	I	II	III	IV	V	4
ORDER OF RUNS(*j*)	1	2	3	4	5	4
OPERATOR (*l*)	α	β	ɣ	δ	ε	4

	ORDER(*j*)					
BATCH (*i*)	1	2	3	4	5	$T_{Batch}(i)$
I	Aα (4)	Bβ (1)	Cɣ(2)	Dδ (2)	Eε(2)	11
II	Bδ(2)	Cε(2)	Dα(3)	Eβ(2)	Aɣ(3)	12
III	Cβ(1)	Dɣ(2)	Eδ(3)	Aε(3)	Bα(3)	12
IV	Dε (3)	Eα(4)	Aβ(3)	Bɣ(3)	Cδ(3)	16
V	Eɣ (2)	Aδ(1)	Bε(3)	Cα(3)	Dβ(2)	11
$T_{Order}(j)$	12	10	14	13	13	62
$T_{Treatment}(k)$	A	B	C	D	E	Total
	14	12	11	12	13	62
$T_{Operator}$	α	β	ɣ	δ	ε	Total
	17	9	11	13	12	62

Table 19 Data for Balanced Incomplete Block

	Operators (Treatment)				
DAYS	**A**	**B**	**C**	**D**	$T_{i\cdot}$
M	**-7**	**-3**	**-5**	**-**	**-15**
T	**10**	**-**	**7**	**9**	**26**
W	**-**	**3**	**3**	**-3**	**3**
R	**-1**	**-7**	**-**	**-3**	**-11**
$T_{j\cdot}$	2	-7	5	3	$T_{\bullet\bullet}$ =3

We will use a new notation to make it easier to do the analysis. Define:

b= number of blocks in the experiment ($b = 4$)

t =number of treatments in the experiment ($t = 4$)

k= number of treatments per block ($k = 3$)

r=number of replications of a given treatment throughout the experiment ($r = 3$)

N= total number of observation = $(b)(k) = 12$

λ= number of times each pair of treatments (say A & B) appear together throughout the experiment

$$\lambda = \frac{r(k-1)}{(t-1)} = \frac{3(2)}{3} = 2 .$$

Check if A & B appear together on Monday and on Thursday. A & C appear together on Monday and Tuesday. A & D appear together on Tuesday and Thursday. You confirm the rest the same way.

We now calculate the Sum of Squares

Note: Correction Factor CF: $CF = \frac{T_{\bullet\bullet}^2}{N} = \frac{(3)^2}{12} = 0.75$

SS Total: $SS_T = \sum\sum y_{ij}^2 - CF \Rightarrow 399 - \frac{(3)^2}{12} = 399 - 0.75 = 398.25$

Unadjusted SS (block): It is uncorrected or unadjusted because we ignored the missing treatments.

Unadjuated SS(Block):

$$SS_B = \sum_{i=1}^{b=4} \frac{T_{i\bullet}^2}{k} - CF = \left(\frac{(-15)^2 + (26)^2 + (3^3) + (-11)^2}{3} \right) - CF = 342.92$$

SS Treatment (adjust for Block), $SS_{Treat} = \frac{\sum_{j=1}^{t=4} Q_j^2}{k\lambda t}$, $where \;\; Q_j = kT_{\bullet j} - \sum_i n_{ij} T_{i\bullet}$

Where n_{ij} = 1 if treatment j appears in block i and 0 if treatment j does not appear in block i.

Note that $\sum_i n_{ij} T_{i\bullet}$, is merely the sum of all block which contain treatment j.

Q_1=3(2)-[- 15+26-11]=6-0=6, Q_2=3(-7)-[-15+3-11] =-21+23=2

Q_2=3(-7)-[-15+3-11] =-21+23=2

Q_3=3(5)-[-15+26+3] =15-(14) =1, Q_4=3(3)- [26+3-11] =9-18= -9

$$\sum Q_j = 0 = (6+2+1-9)$$

The sum of these Q values will be zero because they form a contrast whose sum of squares completely partition the treatment sum of squares, as we saw in the case of orthogonal contrasts. As a result, we did not subtract CF, because it was taken care of during the adjustment for blocks.

$$SS_{Treat} = \frac{6^2 + 2^2 + 1^2 + (-9)^2}{(3)(2(4))} = 5.04$$

$$SS_{Error} = 398.25 - 342.92 - 5.04 = 50.25$$

Note that we cannot test for blocks using the current form because the SS for blocks has not been

unadjusted for due to the missing treatment values. In the case of a symmetric balanced incomplete randomized block design where $t = b$, the block sum of squares may be adjusted just like we adjusted the treatment sum of squares and then we can test for its significance. **However, we still cannot test for both at the same time. We will test for one, ignore the other, and then test for the other**. In our example, since $t = b$, we can test for blocks by repeating the analysis and now SS treatment will be uncorrected. You will notice that the value of SS (treatment) would be larger than what we have now.

Table 20 ANOVA for Balanced Incomplete Design (Treatment Adjusted)

Source	SS	Df	MS	F
Block (Unadjusted)	(342.92)	3	-	
Treat (adjusted)	5.08	3	1.69	<1(n.s.)
Error	50.25	5	10.05	
TOTAL	398.25	11		

For adjusted SS for Blocks: $SS_{Block} = \dfrac{\sum_{i=1}^{b=4}\left(Q_i'\right)^2}{r\lambda b}$, *where* $Q_i' = rT_{i\bullet} - \sum_j n_{ij}T_{\bullet j}$

Q'_1=3(-15)- [2-7+5] = -45+0 =-45
Q'_2=3(26) - [2+5+3] = 78-10 = 68
Q'_3=3(3) - [-7+5+3] = 9-1 = 8
Q'_4=3(-11)- [2-7+3] = -33+2=-31

$$\sum Q_i = 0 = (-45+68+8-31)$$

$$SS_{Block} = \frac{(-45)^2+(68)^2+8^2+(-31)^2}{(3)(2(4))} = 319.75$$

$$Unadjusted: SS_{Treat} = \frac{2^2+(-7)^2+5^2+3^3}{3} - CF = 29 - 0.75 = 28.25$$

$$SS_{Error} = 398.25 - 319.75 - 28.25 = 50.25$$

Table 21 ANOVA for Balanced Incomplete Design (Block Adjusted)

Source	SS	DF	MS	F
Block (adjusted)	319.75	3	106.58	10.6 **
Treat (unadjusted)	(28.25)	3	-	
Error	50.25	5	10.05	
TOTAL	398.25	11		

Significant at α=0.025
$F_{0.025,\,3,5} = 7.76$

SUMMARY

Design of experiments is an activity that every Engineer should take seriously. Engineers are called upon every day to make decisions regarding programs, processes and systems that have significant implications on the safety and well-being of society, be they chemical processes, the environment, infrastructure, machinery and equipment, or others. While Engineers are known for sound and fact based judgment, that may not be enough, especially when they are called upon to make decisions regarding variables and factors with uncertainties that are mostly random and of questionable predictability. Those situations require an understanding of the formal schemes and structures necessary to deal with variability, bias, and randomness.

In this course, we have dealt with some of the common and vexing issues related to the design of experiment under the overarching themes of Planning, Design, and Analysis. The major issues that require emphasis and particular attention regarding good experimental test plans include:

- A true experiment is one in which certain independent variables are manipulated and their levels assigned at random in order to determine their effect on one or more response variables. The notions of Manipulation and Randomization are essential for a true experiment in order to infer cause and effect.
- A major goal of any experiment is to determine the effects of the factors and, more importantly, to ensure that the estimate of pure error is not masked by nuisance factors. We have discussed the need for Blocking and Randomization to reduce the effect of nuisance factors and noise.
- There are experiments, namely, *Quasi-Experiments*, where it is not possible to Randomize**.** For example, in an attempt to determine the effect of two methods of in-plant instruction, it may not be physically possible to randomly assign the workers to the two instruction types because the classes may have been formed by shifts, division or operation. In this case, a flip of the coin can determine which shift or group goes to which method of instruction. It is also important, however, in this type of experiment, to show that the groups are similar.
- Sometimes, the engineer is forced to deal with an '***Ex-Post-Facto'*** situation. An 'ex-post-facto' is a situation where the engineer has little to do with the independent variables. The independent variable would already have been acted upon and the engineer simply examines the results or the effect of the factor (x)on the response (y). As an example, suppose the effect of rainfall and sunshine (independent variables) on the yield of a plot of land is of interest. It is clear that although both (rainfall and sunshine) are independent variables, they are not independent in the pure sense of the word because no manipulation of the variables would be possible, unless of course the engineer happens to be a rainmaker and/or a sun charmer. Thus, inferences about these situations must be made with a lot of care.

- Randomized Complete Block Designs (RCBD) are single factor designs with one or more restrictions on randomization. They are important designs that help to reduce the confounding effect of nuisance factors on the error variance. Randomized Incomplete Blocks are very practical designs that arise from situations that any Engineer may encounter during the process of acquiring data for engineering decisions. We provide a different notation and an approach to assist in solving this practical engineering design problem.

REFERENCES

1. Taguchi , G., Chowdhury, S. , Taguch, S., (2000), Robust Engineering: Learn How to Boost Quality While Reducing Costs & Time to Market 1st Ed, McGraw Hill, New York, NY
2. Taguchi, G., Elsayed, A. E., Hsiang, T.C., (1989), Quality Engineering in Production Systems, McGraw- Hill Series in Industrial Engineering and Management Science, NY
3. Montgomery, D.C., and Runger, G. (2011), Applied Statistics and Probability for Engineers, 5th ed., John Wiley and Sons, NY.
4. Walpole, R., Meyers, R.H. (2002), Probability and Statistics for Engineers and Scientists, 7th ed., Prentice- Hall Inc., Upper Saddle River, NJ.
5. Devore, J.L., (2012), Probability and Statistics for Engineering and the Sciences, 8th ed., Duxbury & Brooks/Cole, Boston, MA.
6. Montgomery, D. (2012), Design &Analysis of Experiments, 8thed., Wiley & Sons, NY.
7. Vardeman, S.(1994), Statistics for engineering problem solving, 1st ed., PWS Pub. Co., Boston MA
8. Hinkelmann, K., Kempthorne, O., (1994), Design and analysis of experiments, Volume I, John Wiley and Sons, NY
9. Petersen, R. G., (1985), Design and analysis of experiments, Marcel Dekker, New York, NY
10. Hicks, C, R., Turner, K. V. (1999), Fundamental Concepts in the Design of Experiments, 5th Ed, Oxford University Press, New York, NY
11. Hicks, C., R. (1963), Fundamental Concepts in the Design of Experiments, 5th Ed, Holt, Rinehart and Winston, New York, NY

