ANALYSIS OF VARIANCE

Hypothesis concerning difference between the means of two populations can be tested using usual *t-test*. But problem of deciding whether observed differences among more than two sample means can be attributed to chance or whether there are real differences among the means of the populations sampled cannot be tackled using *t*-test. In such cases we use Analysis Of Variance (ANOVA) technique to examine the existence of real difference between the sample averages. In ANOVA, outcomes or dependent variables are continuous, whereas explanatory or independent variables are categorical.

It may seem odd to that a procedure that compares means is called analysis of variance. However, this name is derived from the fact that in order to test for statistical significance between means, we are actually comparing (i.e., analyzing) variances. That is, instead of using mean difference we use variance of the group means about the grand mean over all the groups. Clearly, variance of the group means will be large (or, small) if the group means are distinctly different (or, nearly equal).

In many Industrial situations data are classified into various sources, say suppliers, shifts, machines, type of products, etc. In such situation's ANOVA technique is employed with following purposes:

i) **To identify the dominant source(s), from the list of suspects that are responsible for the change in the average level of the corresponding quality characteristic.**
ii) **To isolate and estimate the contribution of different sources towards the total variability in the quality characteristic.**

According to the number of sources of variation involved, there may be one way / many way analyses of variance.

Example 1

Four chemists (Groups) are asked to determine the percentage of methyl alcohol in a certain chemical compound. Each chemist makes three determinations (replications), and the results are the following:

Chemist	Percentage of Methyl Alcohol			Total	Average
1	84.99	84.04	84.38	253.41	84.47
2	85.15	85.13	84.88	255.16	85.05
3	84.72	84.48	85.16	254.36	84.79
4	84.20	84.10	84.55	252.85	84.28

The means of these four samples are 84.47, 85.05, 84.79 and 84.28 and we want to know whether the differences among them are significant or they can be attributed to chance.

ASSUMPTIONS OF ANOVA

1. **Observations in each group is approximately normal.**
   - **check this by looking at histograms and/or normal probability plots, or use assumption, [for small data set, there really isn't a good way to check normality and we make the common assumption that physical measurements tend to follow normal distribution]**
   - **can handle some non-normality, but not severe outliers.**
2. **Standard deviations of each group are approximately equal.**
   - **rule of thumb: ratio of largest to smallest sample standard deviation must be less than 2. [largest < 2 times smallest]**
3. **Observations are independent. [i.e. error terms are uncorrelated]**

THE PARTITIONING OF SUMS OF SQUARES

At the heart of ANOVA is the fact that variances can be divided up, that is, partitioned. Remember that the variance is computed as the sum of squared deviations from the overall mean, divided by *n-1* (sample size minus one). Thus, given a certain *n*, the variance is a function of the sums of (deviation) squares, or *SS* for short.

In ANOVA, the partition is done as:

Total SS	=	Between Groups SS	+	Within Group SS
(based on total *n* observation)		(based on Group averages)		(based on observations within each Group)

$$\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{..})^2 \rightarrow n\sum_{i=1}^{a}(\bar{y}_{i.}-\bar{y}_{..})^2 + \sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{i.})^2$$

Let us consider the following data:

	Group 1	Group 2
Observation 1	2	6
Observation 2	3	7
Observation 3	1	5
Group Mean	2	6
Within Group Sums of Squares (SSE)	2	2
Overall Mean	4	
Total Sums of Squares (TSS)	28	

The means for the two groups is quite different (*2* and *6*, respectively). The sums of squares *within* each group are equal to *2*. Adding them together, we get total within group SS as *4*. If we now repeat these

computations, ignoring group membership, that is, if we compute the total *SS* based on the overall mean, we get the number *28*. In other words, computing the variance (sums of squares) based on the within-group variability yields a much smaller estimate of variance than computing it based on the total variability (the overall mean). The reason for this in the above example is of course that there is a large difference between means, and it is this difference that accounts for the difference in the *SS*. In fact, if we were to perform an ANOVA on the above data, we would get the following result:

	MAIN EFFECT				
	SS	df	MS	F	p
Effect	24.0	1	24.0	24.0	.008
Error	4.0	4	1.0		

As can be seen in the above table, the total *SS (28)* was partitioned into the *SS* due to *within*-group variability (*2+2=4*) and variability due to differences between means (*28-(2+2) =24*).

SS Error and SS Effect.

The within-group variability (*SS Error*) is usually referred to as *Error* variance. This term denotes the variation in response that we cannot readily explain or account for it in the current design. However, the *SS Effect* we can explain. Namely, it is due to the differences in means between the groups. Put another way, group membership explains this variability because we know that it is due to the differences in means.

SS *Effect* is defined as the sum of square of differences of group means about the overall / grand mean. Now, if group means are close to each other SS *Effect* will be small, otherwise if the group means are distinctly different then SS *Effect* will be large.

Replication vs Repetition

Consider an example on the effect of 3 different exercise program on persons with elevated cholesterol levels.

If a person is subjected to an exercise program and at the completion of the program, his cholesterol level was measured 5 times. In this case 5 is the number of repetitions. Difference between these observed values are due to variation in measurement process only and reflects inherent variability in the measurement system.

Now suppose 5 randomly selected persons are subjected to the same exercise program and at the completion of the program their cholesterol levels are measured. This is the case of replication, where the treatment is replicated 5 times. Variation exhibited by such a scheme gives an estimate of Experimental Error, or, Within Group Variation.

ONE WAY Classification

In general, in a population like this, we have *random samples* **of size *n* from** a **different levels of a single independent variable (or, factor). Sometimes, each factor level is called a treatment <***in analogy to the term used in agricultural experiment, where different fertilizers applied to a portion of soil were regarded as different treatments***> .The** j**th value from the** i**th treatment is denoted by** y_{ij}**, that is,**

Treatment	Observations				Total	Average
1	y_{11}	y_{12}	$\cdots$	y_{1n}	$y_{1\cdot}$	$\bar{y}_{1\cdot}$
2	y_{21}	y_{22}	$\cdots$	y_{2n}	$y_{2\cdot}$	$\bar{y}_{2\cdot}$
$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$	$\vdots$	$\vdots$
a	y_{a1}	y_{a2}	$\cdots$	y_{an}	$y_{a\cdot}$	$\bar{y}_{a\cdot}$

The observations above can be represented by the linear statistical model (mean form)

$$y_{ij} = \mu_i + \varepsilon_{ij}, \forall\, i = 1, 2, \cdots, a \text{ and } j = 1, 2, \cdots, n$$

where μ_i is the mean response at i-th level of the factor and $\varepsilon_{ij} \cap N(0, \sigma^2)$.

If the null hypothesis of equal means is true, then above model reduces to

$$y_{ij} = \mu + \varepsilon_{ij}$$

whereas if the null hypothesis is false, sample mean belonging to i-th factor level will deviate from the grand mean μ by an amount, say, τ_i. So, the model in such case will be (effect form)

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij}, \forall\, i = 1, 2, \cdots, a \text{ and } j = 1, 2, \cdots, n\,,$$

where τ_i is a parameter associated with the i-th treatment and is called the treatment effect. Since $\varepsilon_{ij} \cap N(0, \sigma^2)$, $y_{ij} \cap N(\mu_i, \sigma^2)$, where the variance σ^2 is assumed to be constant at all levels of the factor.

It is required that the experiments be performed in random order so that the variations in the experimental units in which the treatments are applied averages out.

As a case in point, consider an experiment, described earlier, designed to determine the effect of three different exercise programs on the cholesterol level of patients with elevated cholesterol. Each patient is referred to as an experimental unit, the response variable is the cholesterol level of the patient at the completion of the program, and the exercise program is the factor whose effect on cholesterol level is being investigated. Each of the three exercise programs is referred to as a treatment. Such an experimental design where treatments are randomly assigned to the experimental units is also known as completely randomized design.

The treatment levels (or, factor levels) in the experiment could have been chosen in two different ways.

1) **The experimenter might choose a fixed treatment value (i.e. levels). Here one can only test the hypothesis about the specific treatment means and cannot extend the conclusion to the treatment values that were not considered. This is called the fixed-effect model. Here the treatment effects τ_i are assumed to be fixed.**

2) **Alternatively, the a treatments might be a random sample from a larger population of treatments. Here one can extend the conclusion, based on the sample of treatments, to all treatments in the population. Here the treatment effects τ_i are random variables, and one is interested about the variability of τ_i and try to estimate that. This is called random-effect model.**

ANOVA for Fixed-effect Model

In fixed-effects model, the treatment effects τ_i are defined as deviations from the overall mean μ, i.e. $\tau_i = \mu_i - \mu$, so that

$$\sum_{i=1}^{a} \tau_i = 0$$

The null hypothesis of $H_0: \mu_1 = \mu_2 = \cdots = \mu_a$ that we want to test is thus equivalent to

$$H_0: \tau_i = 0 \qquad \textit{for all i = 1, 2, …., a.}$$

Consequently, the alternate hypothesis would be

$$H_1: \tau_i \neq 0 \qquad \textit{for at least one value of i.}$$

The test is based on an analysis of total variability of the combined data [($an-1$)-*times their variance*], which is given by

$$\sum_{i=1}^{a}\sum_{j=1}^{n}\left(y_{ij}-\bar{y}_{\bullet\bullet}\right)^2, \quad \text{where} \quad \bar{y}_{\bullet\bullet}=\frac{1}{na}\sum_{i=1}^{a}\sum_{j=n}^{n}y_{ij}\ .$$

If H_0 is true, all this variability is due to chance, otherwise part of the above variation (or, sum of squares) is due to differences among the population means. To separate these two contributions, we divide this total variability as below:

$$\sum_{i=1}^{a}\sum_{j=1}^{n}\left(y_{ij}-\bar{y}_{\bullet\bullet}\right)^2=n.\sum_{i=1}^{a}\left(\bar{y}_{i\bullet}-\bar{y}_{\bullet\bullet}\right)^2+\sum_{i=1}^{a}\sum_{j=1}^{n}\left(y_{ij}-\bar{y}_{i\bullet}\right)^2+2\sum_{i=1}^{a}\sum_{j=1}^{n}\left(\bar{y}_{i\bullet}-\bar{y}_{\bullet\bullet}\right)\left(y_{ij}-\bar{y}_{i\bullet}\right)$$

where $\bar{y}_{i\bullet}$ is the mean of the observation from the ith population and $\bar{y}_{\bullet\bullet}$ is the mean of the all $a \times n$ observation. Since the cross-product term is zero, above expression reduces to

$$\sum_{i=1}^{a}\sum_{j=1}^{n}\left(y_{ij}-\bar{y}_{\bullet\bullet}\right)^2=n.\sum_{i=1}^{a}\left(\bar{y}_{i\bullet}-\bar{y}_{\bullet\bullet}\right)^2+\sum_{i=1}^{a}\sum_{j=1}^{n}\left(y_{ij}-\bar{y}_{i\bullet}\right)^2$$

Total Sum of Squares = Treatment Sum of Squares + Error Sum of Squares

i.e. $\mathbf{TSS = SS_T + SS_E}$

It may be noted in above expression that, if groups means are close to each other than the first term in the right hand side, i.e. SS_T will be small (may even be close to zero) and as a result SS_E will be large compared to SS_T.

Moreover, in above expression SS_E is a measure of the chance variation (i.e. variation within the sample) and SS_T is a measure of a) chance variation when the null hypothesis is true, and b) chance variation & variation among the sample means when the null hypothesis is false.

Since under the null hypothesis $\bar{y}_{i\cdot}(=\mu+\bar{\varepsilon}_{i\cdot})$ are the values of independent random variables having identical normal distribution with mean μ and variance $\frac{\sigma^2}{n}$, it follows that under H_0 [since $\hat{\mu}=\bar{y}_{\cdot\cdot}$]

$$\frac{n}{\sigma^2}\cdot\sum_{i=1}^{a}\left(\bar{y}_{i\bullet}-\bar{y}_{\bullet\bullet}\right)^2 \cap \chi^2_{a-1}, \text{ that is } \frac{1}{\sigma^2}.SS_T \cap \chi^2_{a-1}.$$

It can be shown that, $E(SS_T)=(a-1)\sigma^2+n\sum_{i=1}^{a}\tau_i^2$. Thus, *under null hypothesis*

$$E(SS_T)=(a-1)\sigma^2 \quad \text{or,} \quad E\left[SS_T/(a-1)\right]=\sigma^2.$$

So, under null hypothesis, $SS_T/(a-1)$ is an unbiased estimate of σ^2 and is called Treatment Mean Square (MS_T). Clearly, if null hypothesis is false $E\left(\frac{SS_T}{a-1}\right)=E(MS_T)=\sigma^2+\frac{n}{a-1}\sum_{i=1}^{a}\tau_i^2$.

Since, for each value of i, y_{ij}'s are values of a random sample of size n drawn from normal population with mean μ_i and variance σ^2 {since $E(y_{ij})=E(\mu_i+\varepsilon_{ij})=\mu_i$ and $Var(y_{ij})=Var(\varepsilon_{ij})=\sigma^2$}, so that using $\hat{\mu}_i=\bar{y}_{i\cdot}$ we find

$$\frac{1}{\sigma^2}\cdot\sum_{j=1}^{n}\left(y_{ij}-\bar{y}_{i\bullet}\right)^2 \cap \chi^2 \text{ with } (n\text{-}1) \text{ degrees of freedom.}$$

Furthermore, since the a random samples are independent, it follows that

$$\frac{1}{\sigma^2}\cdot\sum_{i=1}^{a}\sum_{j=1}^{n}\left(y_{ij}-\bar{y}_{i\bullet}\right)^2 \cap \chi^2_{a(n-1)}, \text{ that is } \frac{1}{\sigma^2}.SS_E \cap \chi^2_{a(n-1)}.$$

Moreover, $SS_E/[a(n-1)]$ is an unbiased estimate of σ^2 and is called Error Mean Square (MS_E). This is valid irrespective of whether the null hypothesis is true or false.

We have seen that if the null hypothesis is false, MS_T provided an estimate of σ^2 plus variation due to difference in treatment means, whereas MS_E provides estimate of σ^2 only. This suggests that we reject the null hypothesis if MS_T is appreciably larger than MS_E, that is, MS_T/MS_E is appreciably larger than 1.

Now,

$$\frac{MS_T}{MS_E} = \frac{\frac{SS_T}{(a-1)\sigma^2}}{\frac{SS_E}{a(n-1)\sigma^2}} \cap F_{(a-1),a(n-1)}.$$

Since, ratio of two independent chi-square random variables, each divided by corresponding degrees of freedom, follow F distribution.

Thus we reject the null hypothesis that the populations are equal if the obtained value of F exceeds $F_{\alpha,a-1,a(n-1)}$, where level of significance is α. Calculation summary of one-way analysis of variation is usually presented in following tabular form:

Sources of Variation	*Degrees of Freedom*	*Sum of Squares*	*Mean Square*	*F*
Treatments	***a*-1**	SS_T	$MS_T = SS_T / (a\text{-}1)$	MS_T / MS_E
Error	***a*(*n*-1)**	SS_E	$MS_E = SS_E / a(n\text{-}1)$	
TOTAL	***an*-1**	*TSS*		

For ease of calculation, TSS and SS_T can be simplified, by simple algebra, to:

$$\text{TSS} = \sum_{i=1}^{a}\sum_{j=1}^{n}\left(y_{ij} - \bar{y}_{\bullet\bullet}\right)^2 = \sum_{i=1}^{a}\sum_{j=1}^{n} y_{ij}^2 - CF$$

$$SS_T = \frac{1}{n}\sum_{i=1}^{a} y_{i\bullet}^2 - CF\ ,$$

where, $y_{i\bullet} = \sum_{j=1}^{n} y_{ij}$ and CF = (grand total)²/ an.

The value of SS_E can then be obtained by subtracting SS_T from TSS.

One of the ANOVA assumptions is that all groups have the same standard deviation. In that case, we can estimate the pooled variance as:

$$S_P^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2 + \cdots + (n_a-1)s_a^2}{\sum_{i=1}^{a}(n_i-1)} = \frac{df_1 s_1^2 + df_2 s_2^2 + \cdots df_a s_a^2}{df_1 + df_2 + \cdots + df_2}.$$

Also, we know that

$$s_i^2 = \frac{\sum_{j=1}^{n}(y_{ij}-\bar{y}_{i\cdot})^2}{n_i-1} = \frac{\sum_{j=1}^{n}(y_{ij}-\bar{y}_{i\cdot})^2}{df_i} \Rightarrow df_i s_i^2 = \sum_{j=1}^{n}\left(y_{ij}-\bar{y}_{i\cdot}\right)^2.$$

So, we get

$$\sum_{i=1}^{a} df_i s_i^2 = \sum_{i=1}^{a}\sum_{j=1}^{n}\left(y_{ij}-\bar{y}_{i\bullet}\right)^2 = SS_E$$

$\sum_{i=1}^{a}(n_i-1) = N-a$ [for balanced design, $\sum_{i=1}^{a}(n-1) = a(n-1)$], which is nothing but degrees of freedom for Error (df_E).

Hence, we have

$$s_p^2 = \frac{SS_E}{df_E} = MS_E.$$

Therefore, MS_E is the pooled estimate of the group variances.

Solution to Example 1:

Calculations:

Total number of observations (n) = 12
Sum of all observations (T) = 253.41 + 255.16 + 254.36 + 262.85 = 1015.78
Grand Mean = 1015.78 / 12 = 84.65

Sum of squares of individual observations = $84.99^2 + 84.04^2 + \ldots + 84.55^2$ = 85985.99
Correction factor (CF) = T^2 / n = 1015.78^2 / 12 = 85984.08
Total Sum of Squares (TSS) = 85985.99 – 85984.08 = 1.91

SS_T = = $[253.41^2 + 255.16^2 + 254.36^2 + 252.85^2] / 3$ - CF
= 85985.13 – 85984.08
= 1.05

SS_E= TSS – SS_T = 1.91 – 1.05 = 0.86

ANOVA Table for Four Chemist Example

Sources of Variation	*Degrees of Freedom*	*Sum of Squares*	*Mean Square*	*F value*
Treatments (i.e. Between Chemists)	3	1.05	0.35	3.24
Error (i.e., within Chemist)	8	0.86	0.108	
Total	11	1.91		

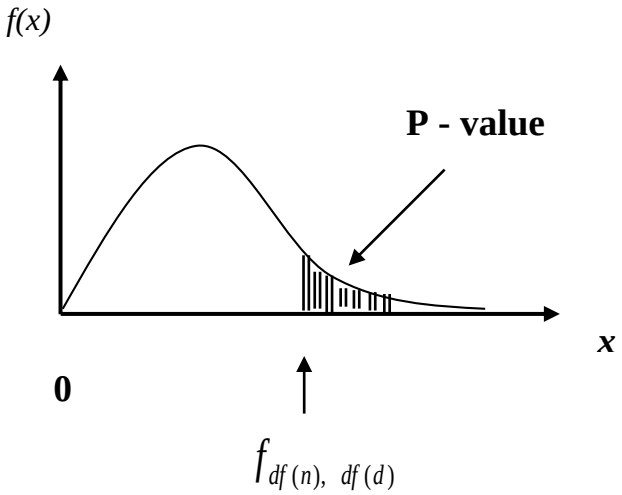
Since computed value of F (3.24) is less than $F_{0.05,3,8}$ (4.07), we cannot reject the null hypothesis and conclude that chemists do not differ significantly in determination of Methyl Alcohol content.

Calculation P-value for F-test

P-value for a statistical test is the smallest level of significance that would lead to rejection of the null hypothesis H_0.

In case of F-test, the P-value of the computed value of the statistic F [$f_{\mathrm{df}(n),\ \mathrm{df}(d)}$] can be obtained by determining the corresponding right tail area. [df(*n*) and df(*d*) are respectively the degrees of freedom of numerator and denominator).

PDF of F distribution with df(*n*) and df(*d*) degrees of freedom



MS Excel Worksheet Function $FDIST(f, \mathrm{df}(n), \mathrm{df}(d))$ gives corresponding P-value.

Confidence Interval for Treatment Means

Frequently, we would like to construct a confidence interval for *i*-th treatment mean. The mean of i-th treatment is

$$\mu_i = \mu + \tau_i, \qquad i = 1,2,\cdots,a$$

A point estimate of μ_i is $\hat{\mu}_i = \bar{y}_{i\cdot}$. Now, since errors are assumed to be normally distributed, each treatment mean $\bar{y}_{i\cdot}(= \mu_i + \bar{\varepsilon}_{i\cdot})$ is normally distributed with mean μ_i and variance σ^2/n. Using MS_E as an estimate of σ^2, we can find the confidence interval based on the *t*-distribution, since

$$t = \frac{\bar{y}_{i\bullet} - \mu_i}{\sqrt{MS_E/n}}$$

follows *t*-distribution with $a(n-1)$ degrees of freedom. Thus the $100(1-\alpha)$% confidence interval on the mean of *i*-th treatment can be defined as

$$\bar{y}_{i\bullet} - t_{\alpha/2,a(n-1)}\sqrt{\frac{MS_E}{n}} \le \mu_i \le \bar{y}_{i\bullet} + t_{\alpha/2,a(n-1)}\sqrt{\frac{MS_E}{n}}$$

Sometimes it may be required to find the confidence interval on the difference in two treatment means, say, $\mu_i - \mu_j \ (i \neq j)$. The point estimator of $\mu_i - \mu_j$ is $\bar{y}_{i\bullet} - \bar{y}_{j\bullet}$ and the variance of the estimator is

$$V(\bar{y}_{i\bullet} - \bar{y}_{j\bullet}) = \frac{\sigma^2}{n} + \frac{\sigma^2}{n} = \frac{2\sigma^2}{n},$$

since the means corresponding to the two treatments are obviously independent.

So,

$$\frac{\left(\bar{y}_{i\bullet}-\bar{y}_{j\bullet}\right)-\left(\mu_i-\mu_j\right)}{\sqrt{2MS_E/n}}, \text{ assuming } \hat{\sigma}^2 = MS_E,$$

has a *t-distribution* with *a(n-1)* degrees of freedom. Therefore, a $100(1-\alpha)\%$ **confidence interval on the difference in two treatment means** $\mu_i - \mu_j$ **is**

$$\bar{y}_{i\bullet}-\bar{y}_{j\bullet}-t_{\alpha/2,\,a(n-1)}\sqrt{\frac{2MS_E}{n}} \le \mu_i-\mu_j \le \bar{y}_{i\bullet}-\bar{y}_{j\bullet}+t_{\alpha/2,\,a(n-1)}\sqrt{\frac{2MS_E}{n}}$$

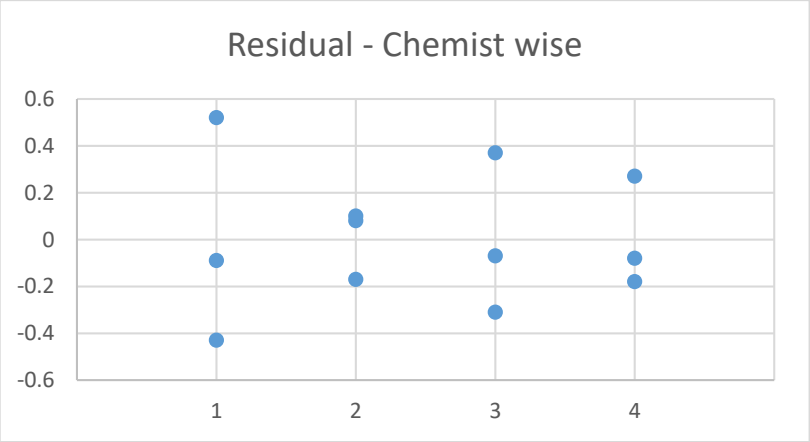
Residual Analysis and Model Checking

The one-way ANOVA assumes that the observations are normally and independently distributed with same variance for each treatment or factor level. These assumptions are checked by examining the residuals. The residuals are defined as $e_{ij} = y_{ij} - \bar{y}_{i\cdot}$ **, which is the difference between the observed value and corresponding treatment mean.**

1. **The normality assumption can be checked by constructing a normal probability plot of the residuals.**

2. **Assumption of equal variance at each factor level can be checked by plotting the residuals against the factor levels and comparing the spread. It is also useful to plot residuals against the fitted value** $\hat{y}_{ij}$**, i.e.** $\bar{y}_{i\bullet}$ **(since for single-factor experiment model** $\hat{y}_{ij} = \bar{y}_{i\bullet}$**); variability in the residuals should be independent of the value of** $\bar{y}_{i\bullet}$ **and should not reveal any obvious pattern. When a pattern appears in these plots some transformation in the original data is called for. For example, if the variability of the residuals increases with** $\bar{y}_{i\bullet}$**, then a transformation such as** $\log y$ *or* $\sqrt{y}$ **should be considered.**

Chemist	Percentage of Methyl Alcohol			Average/ Predicted	Residuals		
1	84.99	84.04	84.38	84.470	0.520	-0.430	0.090
2	85.15	85.13	84.88	85.053	0.097	0.077	-0.173
3	84.72	84.48	85.16	84.787	-0.067	-0.307	0.373
4	84.20	84.10	84.55	84.283	-0.083	-0.183	0.267



3. The independence assumption can be checked by plotting the residuals against time or run order in which the experiment is performed. A pattern in this plot, such as sequence of positive or negative residuals, may indicate that the observations are not independent.

Statistical Test for Equality of Variance

Although residual plots are frequently used to check equality of variance, several statistical tests also have been proposed. These tests may be viewed as formal tests of the hypotheses

$$H_0: \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_a^2$$
$$H_1 : \sigma_i^2 \neq \sigma_j^2, for\ at\ least\ one\ i \neq j$$

One of the widely used tests is Levene's Test. One advantage of Levene's test is that it does not require normality of the underlying data. Levene's test is often used before a comparison of means. We will discuss Brown-Forsythe Test, which is a modified version of Levene's test.

Brown-Forsythe Test uses the absolute deviation of the observations y_{ij} in each treatment from the treatment median, say $\tilde{y}_i$. Absolute deviation based on the median is used as it was found to be more robust against many type of non-normality while retaining good statistical power. Denote these deviations by

$$d_{ij} = |y_{ij} - \tilde{y}_i| \begin{cases} i = 1,2,\cdots,a \\ j = 1,2,\cdots,n \end{cases}.$$

The Brown-Forsythe Test then evaluates whether or not the mean of these deviations (MAD around Median) is equal for all treatments. Considering the analogy with one-way ANOVA F-statistic, he suggested following statistic for testing the equality of variances as

$$W = \frac{n\sum_{i=1}^{a}\left(\bar{d}_{i\bullet} - \bar{d}_{\bullet\bullet}\right)^2 \Big/ (a-1)}{\sum_{i=1}^{a}\sum_{j=1}^{n}\left(d_{ij} - \bar{d}_{i\bullet}\right)^2 \Big/ a(n-1)},$$

which is simply the usual ANOVA *F* statistic for testing the equality means applied to the absolute deviations around median. So, we will perform the usual ANOVA test with d_{ij} as response.

Null hypothesis of equality of treatment variances will be rejected, if the obtained value of $W, say\ w_0$, exceeds $F_{\alpha,\ a-1,a(n-1)}$.

Testing equality of treatment variances for Example 1

Chemist	Percentage of Methyl Alcohol			Median	$d_{ij} = \lvert y_{ij} - \tilde{y}_i \rvert$			Total
1	84.99	84.04	84.38	84.38	0.61	0.34	0	0.95
2	85.15	85.13	84.88	85.13	0.02	0	0.25	0.27
3	84.72	84.48	85.16	84.72	0	0.24	0.44	0.68
4	84.2	84.1	84.55	84.20	0	0.1	0.35	0.45

ANOVA of deviation data, d_{ij}

- **Grand Total = 2.35**
- **Correction Factor = 0.460208**
- **Total SS = 0.474092**
- **SS Chemist = 0.086558**
- **SS Error = 0.387533**

ANOVA Table

Source of variation	Degrees of Freedom	SS	MS	F	F- Crit.
Chemist	3	0.086558	0.028853	0.596	4.07
Error	8	0.387533	0.048442		
Total	11	0.474091			

Therefore, Brown-Forsythe test fails to reject the null hypothesis of equal variances, essentially confirming that the treatment variances are equal.

Bartlett's Test

$$\chi_0^2 = 2.3026\frac{q}{c},$$

where

$$\boldsymbol{q = (N-a)log_{10}S_p^2 - \sum_{i=1}^{n}(n_i-1)log_{10}S_i^2}$$

$$c = 1+\frac{1}{3(a-1)}\left[\sum_{i=1}^{a}\frac{1}{n_i-1}-\frac{1}{N-a}\right]$$

$$S_p^2 = \frac{1}{N-a}\sum_{i=1}^{a}(n_i-1)S_i^2$$ ***[Pooled estimate for the variances]***

It may be noted that S_i^2 is the sample variance of the i-th treatment values. The quantity q is large if the sample variances differ greatly, whereas q is equal to zero if all treatment sample variances are equal. Bartlett (1937) has advocated that the test statistic has approximately a χ^2_{a-1} distribution, therefore, we would reject the null hypothesis if $\chi_0^2 > \chi^2_{\alpha,\ a-1}$.

Bartlett's test is very sensitive to the normality assumption, and therefore it should never be used when the normality assumption is doubtful.

- **Perform Testing equality of treatment variances for Example 1 using Bartlett's Test.**

Test of Individual Treatment Means

Duncan's Multiple Range Test

Suppose the ANOVA indicates that the null hypothesis should be rejected. This implies that there exists difference among the treatment means. But the question is - exactly which treatment means are different? This can be achieved by Duncan's Multiple Range Test (DMRT).

To apply DMRT *for equal sample sizes*, *a* treatment means is arranged in ascending order and the standard error of each mean is determined as

$$SE(\bar{y}_{i\cdot}) = \sqrt{\frac{MS_E}{n}}.$$

For *unequal sample sizes*, replace *n* in above equation by the harmonic mean n_h of the $\{n_i\}$, i.e.

$$n_h = \frac{a}{\sum_{i=1}^{a}(1/n_i)}.$$

In this method, each pair of means is compared against a critical value, called least significant difference, which depends on the difference in ranks of these means in the ordered array.

From Duncan's Table of significant ranges, obtain the values of $r_\alpha(p, df_E)$, for $p = 1, 2, \cdots, a$, where α is the significance level. Convert these ranges into a set of *a*-1 least significant differences using

$$R_p = r_\alpha(p, df_E) \times se(\bar{y}_{i\cdot}), \quad p = 2, 3, \cdots, a$$

Any two means that are "p" distant apart in the ordered arrangement and difference between them is more than R_{p+1}, then those two means are declared to be significantly different.

For example, if $\bar{y}_{(1)}, \bar{y}_{(2)}, \cdots, \bar{y}_{(a)}$ be the ordered arrangement of treatments means, then comparison starts with the largest mean $\bar{y}_{(a)}$

- difference between $\bar{y}_{(a)}$ and $\bar{y}_{(1)}$ is compared with R_a
- difference between $\bar{y}_{(a)}$ and $\bar{y}_{(2)}$ is compared with R_{a-1}

Such comparisons are continued until all means have been compared with the largest mean.

Next the comparison starts w.r.t. to the second largest treatment, $\bar{y}_{(a-1)}$

- difference between $\bar{y}_{(a-1)}$ and $\bar{y}_{(1)}$ is compared with R_{a-1}
- difference between $\bar{y}_{(a-1)}$ and $\bar{y}_{(2)}$ is compared with R_{a-2}

Proceeding similarly difference between $\bar{y}_{(2)}$ and $\bar{y}_{(1)}$ is compared with R_2.

If an observed range is greater than the corresponding least significant difference, then we conclude that the pair of means that are being tested are significantly different.

Difference between a pair of means is not tested if they fall between two other means that do not differ significantly.

Critical Points for Duncan's Multiple Range Test [α = 0.05]

Degrees of freedom ν	p											
	2	3	4	5	6	7	8	9	10	20	50	100
1	18.00	18.00	18.00	18.00	18.00	18.00	18.00	18.00	18.00	18.00	18.00	18.00
2	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09
3	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50
4	3.93	4.01	4.02	4.02	4.02	4.02	4.02	4.02	4.02	4.02	4.02	4.02
5	3.64	3.74	3.79	3.83	3.83	3.83	3.83	3.83	3.83	3.83	3.83	3.83
6	3.46	3.58	3.64	3.68	3.68	3.68	3.68	3.68	3.68	3.68	3.68	3.68
7	3.35	3.47	3.54	3.58	3.60	3.61	3.61	3.61	3.61	3.61	3.61	3.61
8	3.26	3.39	3.47	3.52	3.55	3.56	3.56	3.56	3.56	3.56	3.56	3.56
9	3.20	3.34	3.41	3.47	3.50	3.52	3.52	3.52	3.52	3.52	3.52	3.52
10	3.15	3.30	3.37	3.43	3.46	3.47	3.47	3.47	3.47	3.48	3.48	3.48
11	3.11	3.27	3.35	3.39	3.43	3.44	3.45	3.46	3.46	3.48	3.48	3.48
12	3.08	3.23	3.33	3.36	3.40	3.42	3.44	3.44	3.46	3.48	3.48	3.48
13	3.06	3.21	3.30	3.35	3.38	3.41	3.42	3.44	3.45	3.47	3.47	3.47
14	3.03	3.18	3.27	3.33	3.37	3.39	3.41	3.42	3.44	3.47	3.47	3.47
15	3.01	3.16	3.25	3.31	3.36	3.38	3.40	3.42	3.43	3.47	3.47	3.47
16	3.00	3.15	3.23	3.30	3.34	3.37	3.39	3.41	3.43	3.47	3.47	3.47
17	2.98	3.13	3.22	3.28	3.33	3.36	3.38	3.40	3.42	3.47	3.47	3.47
18	2.97	3.12	3.21	3.27	3.32	3.35	3.37	3.39	3.41	3.47	3.47	3.47
19	2.98	3.11	3.19	3.26	3.31	3.35	3.37	3.39	3.41	3.47	3.47	3.47
20	2.95	3.10	3.18	3.25	3.30	3.34	3.36	3.38	3.40	3.47	3.47	3.47
30	2.89	3.04	3.12	3.20	3.25	3.29	3.32	3.35	3.37	3.47	3.47	3.47
40	2.86	3.01	3.10	3.17	3.22	3.27	3.30	3.33	3.35	3.47	3.47	3.47
60	2.83	2.98	3.08	3.14	3.20	3.24	3.28	3.31	3.33	3.47	3.48	3.48
100	2.80	2.95	3.05	3.12	3.18	3.22	3.26	3.29	3.32	3.47	3.53	3.53
inf	2.77	2.92	3.02	3.09	3.15	3.19	3.23	3.26	3.29	3.47	3.61	3.67

Example 2

The compressive strength of concrete is being and four different mixing techniques are being investigated. Data thus collected given below after coding. <coded value = actual value – 2500>

Mixing Technique	Compressive Strength (psi)			
1	629	500	365	390
2	700	800	475	650
3	300	400	485	550
4	100	200	100	265

Solution >

Mixing Technique	Total	Average	Raw Sum of Squares (RSS*)
1	1884	471.00	930966
2	2625	656.25	1778125
3	1735	433.75	787725
4	665	166.25	130225
TOTAL	6909		3627041

- **Correction Factor (CF) = 2983392.56**
- **TSS = RSS – CF = 3627041 – CF = 643648.44**
- **SS(MT) = 13892531/4 – CF = 489740.19**
- **SS(E) = TSS – SS(MT) = 153908.25**

* Raw SS = $\sum_{j=1}^{4} y_{ij}^2$ for all $1 \leq i \leq 4$.

ANOVA Table

Sources of Variation	Degrees of freedom	Sum of Squares	Mean Square	f_0	P-value
Mixing Technique	3	489740.19	163246.73	12.73	0.00049
Error	12	153908.25	12825.69		
TOTAL	15	643648.44			

Since, P = 0.00049 is considerably smaller than 0.05 (even 0.01), we have strong evidence that treatment means are different.

Let us now apply DMRT to analyze the mean compressive strength of the four mixing techniques.

The treatment means in ascending order are:

Treatment Number	Ordered Arrangement	Mean
4	(1)	166.25
3	(2)	433.75
1	(3)	471.00
2	(4)	656.25

The standard error for each mean is $se(\bar{y}_{i\bullet}) = \sqrt{MS_E/n} = 56.62$. **From, Duncan's table of significant ranges for 12 degrees of freedom with** $\alpha =$ **0.05, we get**

$$r_{0.05}(2,12) = 3.08,\ r_{0.05}(3,12) = 3.23 \text{ and } r_{0.05}(4,12) = 3.33$$

.

Therefore, the least significant differences are

$$R_2 = 3.08 \times 56.62 = 174.39$$
$$R_3 = 3.23 \times 56.62 = 182.88$$
$$R_4 = 3.33 \times 56.62 = 188.54$$

The comparisons among the treatment means are as follows:

(4) vs. (1): Tr. 2 vs. Tr. 4 = 656.25 – 166.25 = 490.00> R_4 (188.54)
(4) vs. (2): Tr. 2 vs. Tr. 3 = 656.25 – 433.75 = 222.50 > R_3 (182.88)
(4) vs. (3): Tr. 2 vs. Tr. 1 = 656.25 – 471.00 = 185.25> R_2 (174.39)

(3) vs. (1): Tr. 1 vs. Tr. 4 = 471.00 – 166.25 = 304.75 > R_3 (182.88)
(3) vs. (2): Tr. 1 vs. Tr. 3 = 471.00 –433.75 = 37.25 < R_2 (174.39)

(2) vs. (1): Tr. 3 vs. Tr. 4 = 433.75 – 166.25 = 267.50 > R_2 (174.39)

From this analysis, we see that there are significant differences between all pairs of means except that of treatments 1 and 3. This implies that mixing techniques 1 and 3 produce approximately the same compressive strength but other mixing techniques produce different compressive strength.

It is often helpful to draw a graph of the treatment means with the means that are not different underlined. This graph clearly reveals the result of the experiment and shows that mixing technique 2 results in maximum compressive strength.

Fisher's Least Significant Difference

The Fisher LSD method compares all pairs of means with the null hypothesis $H_0: \mu_i = \mu_j (\forall\, i \neq j)$ using the *t*-statistic

$$t_0 = \frac{\bar{y}_{i\bullet} - \bar{y}_{j\bullet}}{\sqrt{2MS_E/n}}.$$

Assuming a two-sided alternative hypothesis, the pair of means μ_i and μ_j would be declared significantly different if

$$\frac{|\bar{y}_{i\cdot} - \bar{y}_{j\cdot}|}{\sqrt{2MS_E/n}} > t_{\alpha/2,a(n-1)}$$

$$\text{or, } |\bar{y}_{i\bullet} - \bar{y}_{j\bullet}| > LSD,$$

where LSD, the Least Significant Difference, is

$$LSD = t_{\alpha/2,a(n-1)}\sqrt{\frac{2MS_E}{n}}.$$

If the sample sizes are different in each treatment, the LSD is defined as

$$LSD = t_{\alpha/2,a(n-1)}\sqrt{MS_E\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}.$$

Applying Fisher's LSD method to the previous example (Example 2), we get.

Here LSD is given by

$$LSD = t_{0.025,12}\sqrt{\frac{2MS_E}{n}} = 2.179\sqrt{\frac{2\times 12825.69}{4}} = 174.50$$

Tr. 2 vs. Tr. 4	**= 656.25 - 166.25**	**= 490.00 >174.50**
Tr. 2 vs. Tr. 3	**= 656.25 - 433.75**	**= 222.50 >174.50**
Tr. 2 vs. Tr. 1	**= 656.25 - 471.00**	**= 185.25 >174.50**

Tr. 1 vs. Tr. 4	= 471.00 - 166.25	= 304.75 >174.50
Tr. 1 vs. Tr. 3	= 471.00 - 433.75	= 37.25 <174.50
Tr. 3 vs. Tr. 4	= 433.75 - 166.25	= 267.50 >174.50

So, the final conclusion remained same.

ANOVA for Random Effects Model

The linear statistical model for random effects case is

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij}, \quad i = 1, 2, \cdots, a \text{ and } j = 1, 2, \cdots, n$$

where the treatment effects τ_i and the errors ε_{ij} are independent random variables. It may be noted here that the model is identical in structure to the fixed-effect case, but the parameters have a different interpretation. If the variance of the treatment effects is σ_τ^2, then by virtue of being independence the variance of the response is $Var(y_{ij}) = \sigma_\tau^2 + \sigma^2$.

To test hypotheses in this model, we assume that the errors ε_{ij} are normally and independently distributed with mean 0 and variance σ^2 and the treatment effects τ_i are normally and independently distributed with mean 0 and variance σ_τ^2. Since the treatments (i.e. factor levels) were selected randomly, we are more interested in population of treatments, so we test hypothesis about the variance components

$$H_0: \sigma_\tau^2 = 0$$
$$H_1: \sigma_\tau^2 > 0$$

If $\sigma_\tau^2 = 0$, all treatments are identical in effect, otherwise there exists variation between treatments effects, which necessarily implies that treatments have significant impact on the response.

The ANOVA decomposition of total variability is still valid; that is

$$TSS = SS_T + SS_E$$

and expressions for TSS, SS_T and SS_E remains the same as that of a fixed effect model.

In random effects model, the expected values of the mean squares for treatments and error are given by

$$E(MS_T) = \sigma^2 + n\sigma_\tau^2 \text{ and } E(MS_E) = \sigma^2.$$

It can be seen that, if the null hypothesis is true then both MS_T and MS_E estimates σ^2. Furthermore, both MS_T and MS_E are independent. Consequently, the ratio

$$F_0 = \frac{MS_T}{MS_E}$$

is a *F* random variable with a-1 and $a(n$-1) degrees of freedom under null hypothesis. The null hypothesis would be rejected at the α level of significance if the computed value of the test statistic $f_0 > f_{\alpha,\ a-1,a(n-1)}$.

In random effect model, in addition to testing the significance of differences in treatment effect, we also want to estimate the variance components ($\sigma^2$and σ_τ^2) in the model. The estimators of the variance components are

$$\hat{\sigma}^2 = MS_E \text{ and}$$

$$\hat{\sigma}_\tau^2 = \frac{MS_T - MS_E}{n}.$$

Randomized Complete Block Design

In any experiment, variability arising from a nuisance factor can affect the result. Nuisance factor is defined as a design factor that probably has an effect on the response, but we are not interested in that effect.

Nuisance factor effect can be taken care of by one of the three ways given below:

1. If it is unknown, not measurable and not controllable, we can try to randomize the experiment to balance out its effect.

2. If it is known and measurable but uncontrollable, we can use Analysis of Covariance to compensate its effect.

3. If the nuisance factor is known, measurable and controllable, a design technique called blocking can be used to systematically eliminate its effect on the statistical comparison among treatments.

Suppose an experiment is designed to study the performance of 4 detergents in cleaning clothes having 3 different types of common stains. There is only one factor here – detergent type – and a nuisance factor stain type. Clearly there will be difference between the stains and if we fail to eliminate its effect, this difference will add variability to the experimental error. As a result, the experimental error will reflect both random error and variability between stains. But we would like to remove the variability between strain types from the experimental error.

A design that would accomplish this requires the experimenter to test each detergent randomly once on each of three strain types. This design shown in the next page. The left panel gives the order in which detergents are to be used within a block, while the right panel gives the data (time required to remove the strain) in standard order. Such an experiment is called a randomized complete block design (RCBD).

Experimental Design Matrix

Strain 1	Strain 2	Strain 3
DT 3	DT 2	DT 1
DT 1	DT 1	DT 4
DT 4	DT 3	DT 2
DT 2	DT 4	DT 3

Data Matrix

Detergent	Stain 1	Stain 2	Stain 3
Type 1	45	43	51
Type 2	47	46	52
Type 3	48	50	55
Type 4	42	37	49

It may be noted that in this experiment four white t-shirts were stained with one of three common stains and allowed to sit for a day. The strained shirts, were then washed with all four types of detergent in random order. Same is done for all strain types.

The word complete indicates that each block (strain type) contains all the treatments (detergent type) and the treatments within a block are applied in random order. It may be noted here that each block forms a more homogeneous experimental unit on which one can compare the detergents. Effectively, this design strategy improves the accuracy of comparison among detergent types by eliminating variability among the stain types.

Statistical Analysis of the RCBD

Suppose, we have, '*a*' treatments that are to be compared and '*b*' blocks. The randomized complete block design is shown below. There is one observation per treatment in each block and the order in which the treatments are run within each block is determined randomly. Since the randomization is done within a block only, we say that block represent a restriction on randomization.

Treatment	Block 1	Block 2		Block b	Total	Average
1	y_{11}	y_{12}	$\cdots$	y_{1b}	$y_{1\cdot}$	$\bar{y}_{1\cdot}$
2	y_{21}	y_{22}	$\cdots$	y_{2b}	$y_{2\cdot}$	$\bar{y}_{2\cdot}$
$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$	$\vdots$	$\vdots$
a	y_{a1}	y_{a2}	$\cdots$	y_{ab}	$y_{a\cdot}$	$\bar{y}_{a\cdot}$
Total	$y_{\cdot 1}$	$y_{\cdot 2}$		$y_{\cdot b}$	$y_{\cdot\cdot}$	
Average	$\bar{y}_{\cdot 1}$	$\bar{y}_{\cdot 2}$		$\bar{y}_{\cdot b}$		$\bar{y}_{\cdot\cdot}$

The statistical model for the RCBD can be written as (effects model):

$$y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij} \qquad \begin{cases} i = 1, 2, 3, \cdots, a \\ j = 1, 2, 3, \cdots, b \end{cases}$$

where $\boldsymbol{\mu}$ is the overall mean, $\boldsymbol{\tau_i}$ is the effect of ith treatment, $\boldsymbol{\beta_j}$ is the effect of jth block and $\boldsymbol{\epsilon_{ij}}$ is usual NID$(\mathbf{0}, \boldsymbol{\sigma^2})$ random error term. Remember here there is no k index.

Considering both treatments and blocks to be fixed factor, we can think the treatment and block effects as deviations from the overall mean, so that

$$\sum_{i=1}^{a} \boldsymbol{\tau_i} = \mathbf{0} \qquad \textbf{and} \qquad \sum_{j=1}^{b} \boldsymbol{\beta_j} = \mathbf{0}$$

We also assume that treatments and blocks do not interact, that is, effect of treatment i is the same regardless of which block (or, blocks) it is tested in, i.e. we are assuming that main effects are additive. We are interested in testing the equality of the treatment effects. Since, mean of the ith treatment $\mu_i = \mu + \tau_i$, testing the hypothesis that the treatments are equal, i.e.

$$H_0: \mu_1 = \mu_2 = \cdots = \mu_a$$
$$H_1: \mu_i \neq \mu_j \text{ for at least one } i \neq j$$

is equivalent to

$$H_0 : \tau_1 = \tau_2 = \ldots = \tau_a = 0$$
$$H_1 : \tau_i \neq 0 \text{ at least for one } i$$

Analysis of variance can thus be used in case of CRBD also. Here TSS is partitioned into three components, namely

$$\sum_i \sum_j \left(y_{ij} - \bar{y}_{\bullet\bullet}\right)^2 = b\sum_{i=1}^{a}\left(\bar{y}_{i\bullet} - \bar{y}_{\bullet\bullet}\right)^2 + a\sum_{j=1}^{b}\left(\bar{y}_{\bullet j} - \bar{y}_{\bullet\bullet}\right)^2 + \sum_{i=1}^{a}\sum_{j=1}^{b}\left(y_{ij} - \bar{y}_{i\bullet} - \bar{y}_{\bullet j} + \bar{y}_{\bullet\bullet}\right)^2$$

Above expression is written as $TSS = SS_T + SS_B + SS_E$. Moreover, the degrees of freedom corresponding to above sum of squares are ab-1, (a-1), (b-1) and (a-1)(b-1) respectively. In practice, we compute *TSS*, SS_T *and* SS_B and then obtain SS_E by subtraction.

It can be shown, as earlier, that

$$E\left(SS_T\right) = \left(a-1\right)\sigma^2 + b\sum_{i=1}^{a}\tau_i^2 \qquad E\left(SS_B\right) = \left(b-1\right)\sigma^2 + a\sum_{j=1}^{b}\beta_j^2$$
$$E\left(SS_E\right) = \left(a-1\right)\left(b-1\right)\sigma^2$$

The formulae for calculating sum of squares can be simplified to

$$CF = \frac{y_{\bullet\bullet}^2}{ab}$$
$$TSS = \sum_{i=1}^{a}\sum_{j=1}^{b} y_{ij}^2 - CF$$
$$SS_T = \frac{1}{b}\sum_{i=1}^{a} y_{i\bullet}^2 - CF, \qquad MS_T = \frac{SS_T}{a-1}$$
$$SS_B = \frac{1}{a}\sum_{j=1}^{b} y_{\bullet j}^2 - CF, \qquad MS_B = \frac{SS_B}{b-1}$$
$$SS_E = TSS - SS_T - SS_B \qquad MS_E = \frac{SS_E}{(a-1)(b-1)}$$

If the null hypothesis is true, then MS_T is an unbiased estimate of σ^2, while if null hypothesis false, MS_T estimates σ^2 plus effect of treatments. On the other hand MS_E is always an unbiased estimate of σ^2. To test the null hypothesis that the treatment effects are all zero, we compute

$$F_0 = \frac{MS_T}{MS_E}, \text{ which follows } F_{a-1,(a-1)(b-1)}, \text{ under null hypothesis.}$$

We would reject the null hypothesis at α-level of significance if the computed value of F_0, that is, f_0 conforms to this identity:

$$f_0 > F_{\alpha,a-1,(a-1)(b-1)}$$

These computations are represented in ANOVA table, as given below.

Source of Variation	Degrees of freedom	Sum of Squares	Mean Square	F_0
Treatments	a-1	SS_T	MS_T	MS_T / MS_E
Blocks	b-1	SS_B	MS_B	
Error	(a-1)(b-1)	SS_E (by subtraction)	MS_E	
TOTAL	ab-1	TSS		

One may also be interested in comparing block means because, if these means do not differ greatly, blocking may not be necessary in future experiments. But since blocks represent a restriction on randomization, testing of block means using F-statistic is not advisable and hence the same is not included in the analysis of variance table.

Cleanliness Reading Data – RCBD

Detergent Type	Stain Type			Treatment Totals $y_{i\bullet}$
	1	2	3	
1	45	43	51	139
2	47	46	52	145
3	48	50	55	153
4	42	37	49	128
Block Totals $y_{\bullet j}$	182	176	207	565 $[y_{\bullet\bullet}]$

The sum of squares for ANOVA are computed as follows

CF = $(565)^2 / (4\times3) = 26602.08$

TSS = $\sum_{i=1}^{4}\sum_{j=1}^{3} y_{ij}^2 - CF$ **= $(45)^2 + (43)^2 + \ldots + (49)^2 - 26606.08 = 264.92$**

SS_T = $\frac{1}{3}\sum_{i=1}^{4} y_{i\cdot}^2 - CF$ **= 1/3 $[(139)^2 + \ldots + (128)^2] - 26606.08 = 110.92$**

SS_B = $\frac{1}{4}\sum_{j=1}^{3} y_{\cdot j}^2 - CF$ **= 1/4 $[(182)^2 + \ldots + (207)^2] - 26606.08 = 135.17$**

$SS_E = TSS - SS_T - SS_B = 264.92 - 110.92 - 135.17 = 18.83$

Source of Variation	Degrees of freedom	Sum of Squares	Mean Square	F_0
Detergents (*Treatments*)	3	110.92	36.97	11.77
Stain Samples (*Blocks*)	2	135.17	67.58	
Error	6	18.83	3.14	
TOTAL	11	264.92		

Since $f_0 = 11.77 > F_{0.05,\,3,\,6} = 4.76$, we conclude that there is a significant difference in the different detergents so far as their performance in stain removal is concerned.

TEST OF INDIVIDUAL TREATMENT MEANS

- ***For DMRT,*** $SE(\bar{y}_{i\cdot}) = \sqrt{\frac{MS_E}{b}} = \sqrt{\frac{MS_E}{3}}$
- ***For Fisher's LSD,*** $LSD = t_{\alpha/2,6}\sqrt{\frac{2MS_E}{b}} = t_{\alpha/2,6}\sqrt{\frac{2MS_E}{3}}$

RESIDUAL ANALYSIS AND MODEL CHECKING

Fitted values from the statistical model for RCBD can be shown to be as follow:

$$\hat{y}_{ij} = \bar{y}_{i\cdot} + \bar{y}_{\cdot j} - \bar{y}_{\cdot\cdot}$$

Residuals for the RCBD are just the difference between the observed and estimated (or, fitted) values from the statistical model, that is

$$e_{ij} = y_{ij} - \hat{y}_{ij} = y_{ij} - \bar{y}_{i\bullet} - \bar{y}_{\bullet j} + \bar{y}_{\bullet\bullet}$$

Plots as described in the *residual analysis of one-way design* may be carried out in case of RCBD too.

Repeated Measure One Way ANOVA

In social and behavioral sciences, the experimental units are often individuals (also known as subjects). Because of differences in experience; training; or background, the differences in responses of different individual to the same treatment may be very large in some experimental situations. Unless it is controlled, the variability between subjects would become part of experimental error, and in some cases, would significantly inflate EMS, making it more difficult to detect real differences between treatments.

If different subject groups (i.e. experimental units/individuals) are used for each level of the Independent Variable a one-way independent (i.e. between subjects) ANOVA should be used. But when same group of subjects are tested for each levels of the Independent Variable Repeated Measure One-way ANOVA should be used.

The one-way ANOVA assumes that the treatments are independent, and the repeated measures ANOVA assumes that the treatments are dependent and as such violates the usual ANOVA assumption of independence of treatments.

As with any ANOVA, repeated measures ANOVA tests the equality of means. However, repeated measures ANOVA is used when all members of a random sample are measured under a number of different conditions or at different time points. As the sample is exposed to each condition, the measurement of the dependent variable is repeated.

Thus we can analysis data using repeated measures ANOVA for two types of study design. Studies that investigate either

(1) Changes in mean scores over three or more time points, or

(2) Differences in mean scores under three or more different conditions.

For example, for (1), one might be investigating the effect of a 6-month exercise training program on blood pressure and want to measure blood pressure at 3 separate time points (pre-, midway and post-exercise intervention), which would allow him/her to develop a time-course for any exercise effect. So in this case, the independent variable is Time and treatment levels are different time-points. (Longitudinal data)

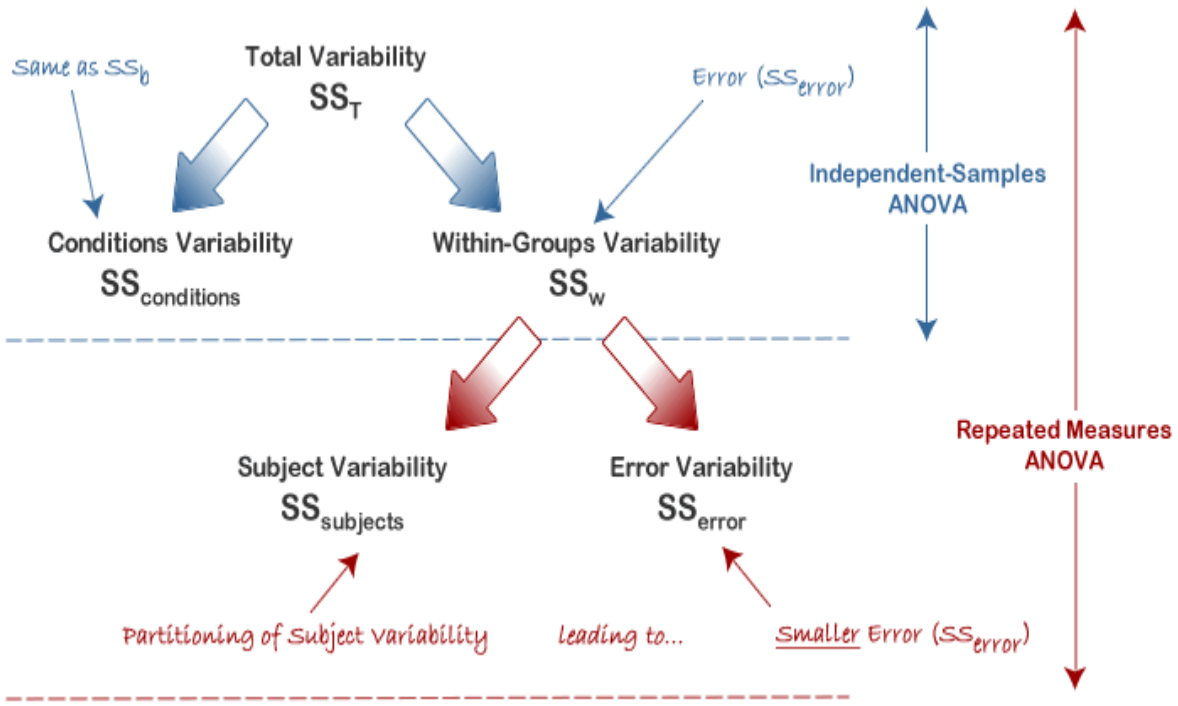
For (2), one might get the same subjects to eat different types of cake (chocolate, caramel and lemon) and rate each one for taste, rather than having different group of people eat different cake flavors. In this case, the independent variable is Type of Cake and its levels are different cake flavors.

The important point with these two study designs is that the same people/subjects are measured repeatedly on the same dependent variable (and hence called repeated measures).

Suppose that an experiment involves "*a*" treatments and every treatment is used exactly once on each of the same "*n*" subjects. The data would appear as:

Treatment (time/condition)	Subject				**Treatment Total**
	1	2	….	*n*	
1	y_{11}	y_{12}	….	y_{1n}	$\boldsymbol{y_{1\cdot}}$
2	y_{21}	y_{22}	….	y_{2n}	$\boldsymbol{y_{2\cdot}}$
. . .	. . .	. . .		. . .	
a	y_{a1}	y_{a2}	….	y_{an}	$\boldsymbol{y_{a\cdot}}$
Subject Total	$\boldsymbol{y_{\cdot 1}}$	$\boldsymbol{y_{\cdot 2}}$		$\boldsymbol{y_{\cdot n}}$	$\boldsymbol{y_{\cdot\cdot}}$

Logic of Repeated Measure ANOVA



The model that is to be used for such design is

$$y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij},$$

where τ_i is the effect of *i*-th treatment and β_j is a parameter associated with the *j*-th subject. We assume that treatments are fixed and subjects employed are a random sample of subjects from a larger population of potential subjects.

Hypothesis for Repeated Measures ANOVA

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_a$$
$$H_1 : \mu_i \neq \mu_j \quad for\ at\ least\ one\ i \neq j$$

For our exercise-training example, the null hypothesis (H_0) is that mean blood pressure is the same at all the time points (pre-, 3 months, and 6 months). The alternative hypothesis is that mean blood pressure is significantly different at one or more time points.

As usual, we partition total sum of squares, as

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

So, we have, $TSS = SS_{\text{Between Treatments}} + SS_{\text{Within Treatments}}$ **and corresponding degrees freedom identity is** $an\text{-}1 = (a\text{-}1) + a(n\text{-}1)$**.**

Now, $SS_{\text{Within Treatments}}$ **depends on**

- **i)** *different in subject effects*, **and**
- **ii)** *uncontrolled variability* (*error* or *noise*)**.**

Therefore, we decompose the sum of squares resulting from differences within treatments as follows:

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

i.e. $SS_{\text{Within Treatments}} = SS_{\text{Subjects}} + SS_{\text{Error}}$ **with corresponding identity for degrees of freedom as**

$$a(n-1) = (n-1) + (a-1)(n-1)\,.$$

To test the hypothesis of no treatment effect, in repeated measure ANOVA, we would use the ratio

$$F_0 = \frac{SS_{Treatments}/(a-1)}{SS_E/(a-1)(n-1)} = \frac{MS_{Treatment}}{MS_E}\,.$$

The null hypothesis would be rejected, if $F_0 > F_{\alpha,a-1,(a-1)(n-1)}$ **.**

The ANOVA procedure for repeated measure are summarized in the next page.

Computing Formulae:

- $TSS = \sum_{i=1}^{a}\sum_{j=1}^{n} y_{ij}^2 - \frac{y_{\bullet\bullet}^2}{an}$
- $SS_{Treatments} = \sum_{i=1}^{a} \frac{y_{i\bullet}^2}{n} - \frac{y_{\bullet\bullet}^2}{an}$
- $SS_{Within\ Treatments} = TSS - SS_{Treatments}$
- $SS_{Between\ Subjects} = \sum_{j=1}^{n} \frac{y_{\bullet j}^2}{a} - \frac{y_{\bullet\bullet}^2}{an}$
- $SS_{Error} = SS_{Within\ Treatments} - SS_{Between\ Subjects}$

ANOVA Table for Repeated Measure Design (Single Factor)

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0
Between Treatments	$SS_{Treatments}$	a - 1	$MS_{Treatments}$	$\frac{MS_{Treatments}}{MS_{Error}}$
Within Treatments	$SS_{Within\ Treatments}$	$a(n$ - 1)		
Subjects	$SS_{Subjects}$	n - 1		
Error	SS_{Error}	$(a-1)(n-1)$	MS_{Error}	
Total	SS_{Total}	**an - 1**		

Exercise>

Following is the example of a 6-month exercise-training intervention where six subjects had their fitness level measured on three occasions: pre-, 3 months, and post-intervention. Their data is shown below along with some initial calculations:

	Exercise Intervention		
Subjects	**Pre-**	**3 Months**	**6 Months**
1	45	50	55
2	42	42	45
3	36	41	43
4	39	35	40
5	51	55	59
6	44	49	56
Monthly Means:	**42.8**	**45.3**	**49.7**

Does exercise affect the fitness levels? Use 5% level of significance.

TWO-WAY Classification

Let an experiment is to be carried out involving two independent variables, sometimes called row and column factors. Each independent variable, or factor, is made up of, or defined by, two or more elements called levels.

Let there are a levels of row factor (A) and b levels of column factor (B). The experiment has n replicates, and each replicate contains all "ab" treatment combinations. The observation in the ij-th cell for the kth replicate is denoted by y_{ijk}. While running the experiment, the "abn" experimental combination will have to be run in random order.

In two-way experimentation, in addition to investigating how different levels of the two independent variables, i.e. row and column factors, affect the dependent variable, we can also test whether levels of one independent variable affect the dependent variable in the same way across the levels of the other independent variable (or vice versa). If the effects are not same, we say there is an interaction between the independent variables. Thus, with two-way experimentation, we can study the effects of the individual independent variables, called the main effects, as well as the interaction effects.

In a two-way design, the effect of varying the levels of the two factors affecting the process output is investigated. Each complete trial or replication of the experiment takes into account all the possible combinations of the varying levels of both the factors. Such a design ensures that the least number of experiment runs are conducted to generate the maximum amount of information about how input variables affect the output of a process.

Meaning of Main Effects

The amount of change, on the average, in the process output for a change in the 'level' of a given factor is referred to as the 'main effect' of that factor. Table 1 shows an example of a two-way experiment involving two factors with two levels each. The two levels of each factor may be denoted as 'low' and 'high', which are usually symbolized by '-' and '+' in factorial designs, respectively.

Table 1: A Simple Two-way Experiment

	A (-)	A(+)
B(-)	20	40
B(+)	30	52

The main effect of a factor is basically the difference in 'average' response as that factor goes from '-' to '+'. Mathematically, this is the difference of two numbers: 1) the average response when the factor is at high level, and 2) the average response when the factor is at low level.

The main effect of A is the difference between the average response when A is (+) and the average response when A is at (-). Now average response when A is at (+) is (40+52)/2=46, whereas the same when A is at (-) is (20+30)/2=25. The main effect of A, therefore, is equal to (46-25) =21.

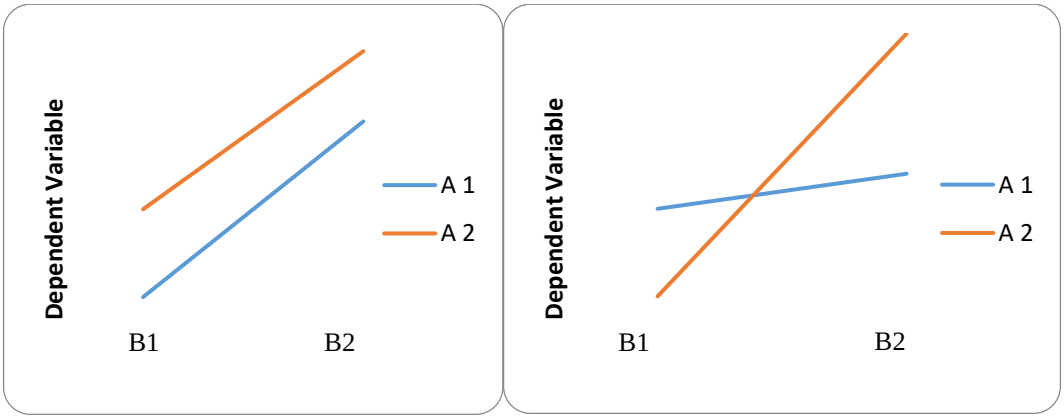
Similarly, the main effect of B is the difference between (52+30)/2 and (20+40)/2, i.e. 11. Here, one can see that the factor A exerts a greater influence on the output of process, having a main effect of 21 versus factor B's main effect of only 11.

Meaning of Interaction

An interaction between the two factors is present in a two-way ANOVA when the *effect of one factor is not the same across the levels of the other factor (or vice versa)*. For two-way ANOVA, the most important research question is - whether there exists a statistically significant interaction between the two factors involved in the study. The interaction deals with the cell means; not main effect means. Existence of interaction implies that the difference between the level-wise means of the first factor is not same for different levels of the second factor.

One *procedure for examining an interaction is to plot the cell means. The value of the dependent variable is placed on the vertical (Y) axis; the level of one of the independent variables is placed on the horizontal (X) axis, and the second independent variable is represented by straight lines.* The decision about which variable to be placed on the *X* axis is arbitrary. Regardless of which variable is placed on this axis, the interpretation of the interaction is, for all practical purposes, the same.

If there is no significant interaction between the two independent variables, the lines connecting the cell means will be parallel or nearly parallel (within sampling fluctuation). The plot of a significant interaction (the lines are not parallel) can have many patterns.



No Interaction **With Interaction**

Example of Interaction: Interaction between adding carbon to steel and quenching. Neither of the two individually has much effect on strength but a combination of the two has a dramatic effect.

An Example

For example, if one is interested in the effects of practice and stress level on memory task performance, one might decide to employ a factorial design. One can manipulate practice by having participants read a list of words either once or five times and stress level by having two conditions: in one (low stress), participants are told that the number of words that they recall is unimportant and in the other (high stress), participants are told that most people can recall almost *all* words in the list, and that they are expected to do so as well. The dependent variable is the number of words recalled from the 30-word list.

Let 4 groups of 25 participants are randomly selected from a population of age group 30-40, and they are assigned to one of these four cells at random. Let's the experiment results in following data and we are to see if there is an interaction in this study.

Mean Number of Words Recalled by Group

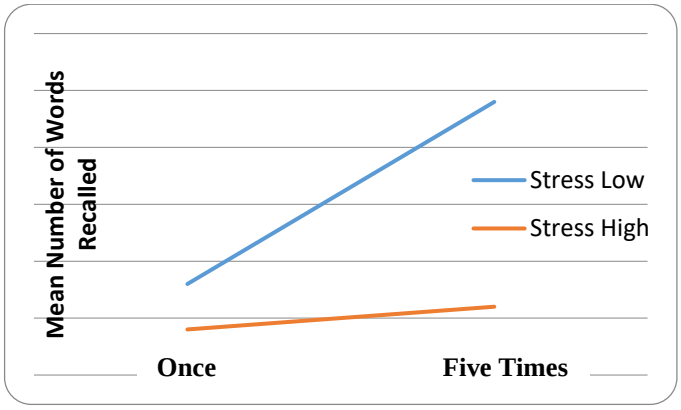
Stress \ Practice	Once	Five Times	
Low	8	24	16
High	4	6	5
	6	15	10.5

The table above indicates the cell means, as well as the marginal means and the grand mean, for the study. For example, the mean number of words recalled under the low stress, one practice condition is 8. This is a cell mean. However, the mean number of words recalled under all low stress conditions (regardless of practice) is 16. This is a marginal mean.

So, do we have evidence of an interaction in this study?

a. ***It appears that there may be a main effect of stress. High stress conditions result in recall of fewer words than low stress conditions.***
b. ***It also appears that there is a main effect of practice: five practices result in better recall of the words than just one practice.***
c. ***However, the effect of the practice variable depends on the level of stress: under low stress conditions, practice seems to have a substantial positive effect (an average of 8 words recalled with one practice and 24 words recalled with five practices), but under high stress conditions, practice has only a small effect (4 versus 6 words under the two practice conditions, respectively).***

Therefore, we have evidence of an interaction in this study. Of course, one will need to carry out the appropriate statistical test before concluding that the evidence is strong enough to support the claim that there is an interaction in the population.



Fixed-Effects Model – Two-way Classification

The observations may be represented using the linear statistical model

$$y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \varepsilon_{ijk} \qquad \begin{cases} i = 1, 2, \ldots, a \\ j = 1, 2, \ldots, b \\ k = 1, 2, \ldots, n \end{cases}$$

where μ **= overall mean,** τ_i **= effect of *i*th level of row factor (*A*),** β_j **= effect of *j*th level of column factor (*B*),** $(\tau\beta)_{ij}$ **= interaction effect between *A and B*, and** ε_{ijk} **is random error with** $NID(0, \sigma^2)$**. In this model, all the factor effects *A*, *B* and *AB* are fixed effects.**

We are interested in testing the hypothesis of no main effects for row factor, no main effect for column factor, and no interaction effect. Since there are two factors in the experiment, the test procedure is sometimes called the *Two - Way ANOVA*. The hypotheses to be tested here are.

$$H_0 : \tau_1 = \tau_2 = \cdots = \tau_a = 0$$
$$H_1 : \text{At least one } \tau_i \neq 0$$

$$H_0 : \beta_1 = \beta_2 = \cdots = \beta_b = 0$$
$$H_1 : \text{At least one } \beta_j \neq 0$$

$$H_0 : (\tau\beta)_{ij} = 0, for\ all\ i, j$$
$$H_1 : \text{At least one } (\tau\beta)_{ij} \neq 0$$

Moreover, we have

$$\sum_{i=1}^{a} \tau_i = \sum_{j=1}^{b} \beta_j = \sum_{i=1}^{a} (\tau\beta)_{ij} = \sum_{j=1}^{b} (\tau\beta)_{ij} = 0$$

Data Structure for Two-way Classification

Factor A	Factor B				Factor A: Average
	1	2	….	*b*	
1	$y_{111}, y_{112}, \cdots, y_{11n}$ $[\bar{y}_{11\cdot}]$	$y_{121}, y_{122}, \cdots, y_{12n}$ $[\bar{y}_{12\cdot}]$	….	$y_{1b1}, y_{1b2}, \cdots, y_{1bn}$ $[\bar{y}_{1b\cdot}]$	$\bar{y}_{1\cdot\cdot}$
2	$y_{211}, y_{212}, \cdots, y_{21n}$ $[\bar{y}_{21\cdot}]$	$y_{221}, y_{222}, \cdots, y_{22n}$ $[\bar{y}_{22\cdot}]$	….	$y_{2b1}, y_{2b2}, \cdots, y_{2bn}$ $[\bar{y}_{2b\cdot}]$	$\bar{y}_{2\cdot\cdot}$
. . .	. . .	. . .		. . .	
a	$y_{a11}, y_{a12}, \cdots, y_{a1n}$ $[\bar{y}_{a1\cdot}]$	$y_{a21}, y_{a22}, \cdots, y_{a2n}$ $[\bar{y}_{a2\cdot}]$	….	$y_{ab1}, y_{ab2}, \cdots, y_{abn}$ $[\bar{y}_{ab\cdot}]$	$\bar{y}_{a\cdot\cdot}$
Factor B: Average	$\bar{y}_{\cdot 1\cdot}$	$\bar{y}_{\cdot 2\cdot}$		$\bar{y}_{\cdot b\cdot}$	$\bar{y}_{\cdots}$

Since we are interested to test the hypothesis about the main factor effects A and B and the AB interaction, as earlier we will decompose total variability in the data into three components, one for each of A, B and AB, and test each component against the error.

After decomposing the total variability and simplifying, we get the following identity

$$\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(y_{ijk}-\bar{y}_{\bullet\bullet\bullet})^2 = bn\sum_{i=1}^{a}(\bar{y}_{i\bullet\bullet}-\bar{y}_{\bullet\bullet\bullet})^2 + an\sum_{j=1}^{b}(\bar{y}_{\bullet j\bullet}-\bar{y}_{\bullet\bullet\bullet})^2$$

$$+ n\sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{y}_{ij\bullet}-\bar{y}_{i\bullet\bullet}-\bar{y}_{\bullet j\bullet}+\bar{y}_{\bullet\bullet\bullet})^2 + \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(y_{ijk}-\bar{y}_{ij\bullet})^2$$

or, $$TSS = SS_A + SS_B + SS_{AB} + SS_E$$

where

$$CF = \frac{y_{\bullet\bullet\bullet}^2}{abn}$$

$$TSS = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n} y_{ijk}^2 - CF$$

$$SS_A = \frac{1}{bn}\sum_{i=1}^{a} y_{i\bullet\bullet}^2 - CF \qquad MS_A = \frac{SS_A}{a-1}$$

$$SS_B = \frac{1}{an}\sum_{j=1}^{b} y_{\bullet j\bullet}^2 - CF \qquad MS_B = \frac{SS_B}{b-1}$$

$$SS_{Interaction} = \frac{1}{n}\sum_{i=1}^{a}\sum_{j=1}^{b} y_{ij\bullet}^2 - CF - SS_A - SS_B \qquad MS_{Interaction} = \frac{SS_{Interaction}}{(a-1)(b-1)}$$

$$SS_E = TSS - SS_A - SS_B - SS_{AB} \qquad MS_E = \frac{SS_E}{ab(n-1)}$$

Since the factors *A* and *B* are fixed factors, it is easy to show that the expected values of these mean squares are

$$\boldsymbol{E(MS_A) = \sigma^2 + \frac{bn\sum_{i=1}^{a}\tau_i^2}{a-1}} \qquad \boldsymbol{E(MS_B) = \sigma^2 + \frac{an\sum_{j=1}^{b}\beta_j^2}{b-1}}$$

$$\boldsymbol{E(MS_{A\times B}) = \sigma^2 + \frac{n\sum_{i=1}^{a}\sum_{j=1}^{b}(\tau\beta)_{ij}^2}{(a-1)(b-1)}} \qquad \boldsymbol{E(MS_E) = \sigma^2}$$

It can be seen that if the null hypothesis about main effects, i.e. $H_0: \tau_i = 0$, $H_0: \beta_j = 0$ and the interaction, i.e. $H_0: (\tau\beta)_{ij} = 0$ are all true, then all four mean squares are unbiased estimates of σ^2.

ANOVA table in such a situation can be represented in the following form

Sources	DF	SS	MS	F_0	Reject H_0 if
A	$a-1$	SS_A	MS_A	$\frac{MS_A}{MS_E}$	$f_0 > F_{\alpha,a-1,ab(n-1)}$
B	$b-1$	SS_B	MS_B	$\frac{MS_B}{MS_E}$	$f_0 > F_{\alpha,b-1,ab(n-1)}$
Interaction (AB)	$(a-1)(b-1)$	SS_{AB}	MS_{AB}	$\frac{MS_{AB}}{MS_E}$	$f_0 > F_{\alpha,(a-1)(b-1),ab(n-1)}$
Error	$ab(n-1)$	SS_E	MS_E		
TOTAL	$abn-1$	TSS			

Example 4

Let us consider the situation of investigating the effect of 4 assembly methods and 3 operators on assembly time of certain product. The data on assembly times (in minutes) are given in the table below. We will use α = 0.01.

Method→ Operator↓	1	2	3	4	Total $(y_{i\cdot\cdot})$
1	38,40 (78)	27,34 (61)	39,36 (75)	44,46 (90)	304
2	42,40 (82)	28,22 (50)	35,40 (75)	38,47 (85)	292
3	41,33 (74)	38,40 (78)	26,30 (56)	44,46 (90)	298
Total $(y_{\cdot j\cdot})$	234	189	206	265	894

In this example, data are classified according to two sources, namely methods and operators, and thus there are 4 sources of variation:

i) Variation between Operators,
ii) Variation between Methods,
iii) Interaction between Operators and Methods, and
iv) Residual variation.

$$y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \varepsilon_{ijk} \quad \begin{cases} i = 1,2,3 \\ j = 1,2,3,4 \\ k = 1,2 \end{cases}$$

Computations:

Total number of observations (n) = 24

Sum of all observations ($y_{...}$) = 894, **CF = $\frac{y_{...}^2}{n}$ = 33301.5**

Total Sum of Squares (TSS) = $38^2 + 40^2 + \ldots + 46^2$ - CF
= 34350 – 33301.5 = 1048.5
Corresponding degree of freedom = n-1 = 24-1 = 23

SS due to operators = $[304^2 + 292^2 + 298^2] / 8$ – CF
= 33310.5 – 33301.5 = 9.0
Corresponding degree of freedom = 3-1 = 2

SS due to method = $[234^2 + 189^2 + 206^2 + 265^2] / 6$ – CF
= 33856.3 – 33301.5 = 554.8
Corresponding degrees of freedom = 4-1 = 3

SS due to interaction between method and operator
= $[78^2 + 61^2 + \ldots + 90^2] / 2$– CF – SS due to method – SS due to operator
= 34200 – 33301.5 – 554.8 – 9.0 = 334.7
Corresponding degrees of freedom = 3 × 2 = 6

Error Sum of Squares = TSS – SS due to operator – SS due to method
- SS due to interaction
= 1048.5 – 9.0 – 554.8 – 334.7 = 150.0
Corresponding degrees freedom = 23 – 2 – 3 - 6 = 12

Test Criteria and Decision-Making

To test the effect of *interaction*, the test statistic is $F = MS_{M\times O} / MS_E = 4.46$. This is tested with the tabulated value of $F_{6,12}$ at 1% level of significance, which is observed to be 4.82. Now since the calculated value of F is less than the tabulated value, we conclude that the *interaction* is not significant.

Now to test the effect of *Method*, the test statistic is $F = MS_M / MS_E = 184.93/12.50 = 14.79$, which much more than the corresponding tabulated value of $F_{3,12}$ at 1% level of significance (5.95). So, we come to the conclusion that *Method* has significant contribution towards total variation in assembly time.

And finally, to test the effect of *Operator*, the test statistic is $F = MS_O / MS_E$ and above value of F is obtained to be less than 1. And any effect with F-value less than 1 will necessarily have insignificant contribution, since for an effect to have certain contribution *F-value* must be greater than 1.

ANOVA Table for Assembly Time data

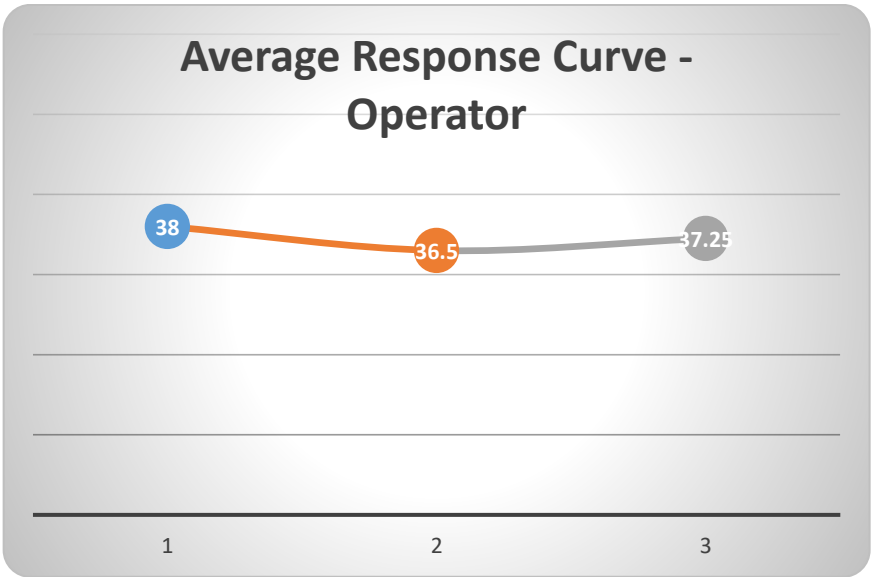
Sources of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F-value	Remark (at 1% level)
Operator (O)	2	9.0	4.50	< 1	Insignificant
Method (M)	3	554.8	184.93	14.79	Significant
O × M	6	334.7	55.78	4.46	Insignificant
Error	12	150.0	12.50		
TOTAL	23	1048.5			

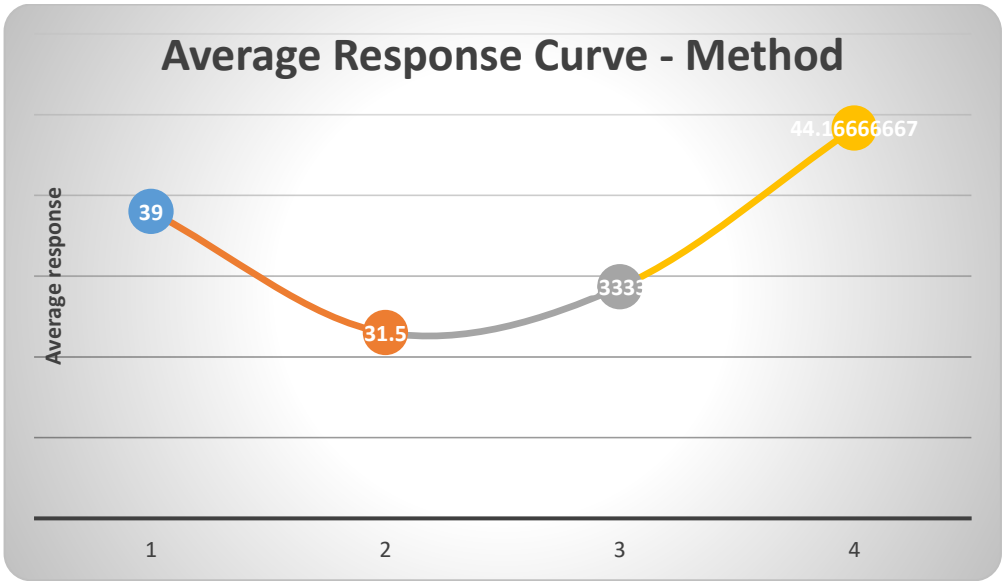
Revised ANOVA for Assembly Time data

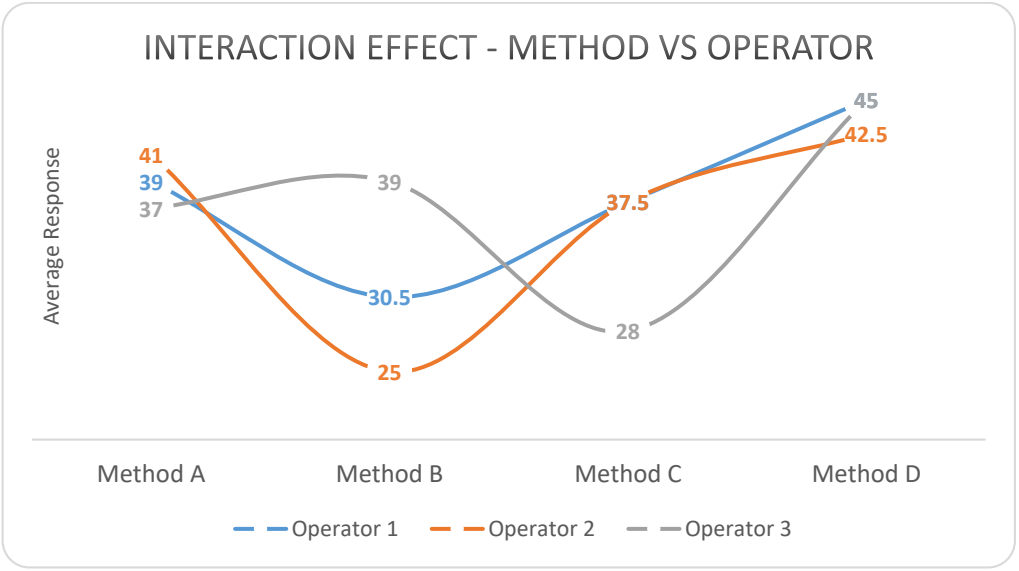
Sources of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F-value	Remark (at 1% level)
Operator (O)	2	9.0	4.50	< 1	Insignificant
Method (M)	3	554.8	184.93	6.87	Significant
Error (Pooled)	18	484.7	26.93		
TOTAL	23	1048.5			

Graphical Representation of Effects

Method → Operator ↓	*Method* 1	*Method* 2	*Method* 3	*Method* 4	Operator Average
Operator 1	39	30.5	37.5	45	38
Operator 2	41	25	37.5	42.5	36.5
Operator 3	37	39	28	45	37.25
Method Average	39	31.5	34.33	44.17	37.25




Average Response Curve - Method
Average response
39
31.5
44.16666667
1
2
3
4


INTERACTION EFFECT - METHOD VS OPERATOR
Average Response
41
39
37
39
30.5
25
37.5
45
42.5
28
Method A
Method B
Method C
Method D
Operator 1
Operator 2
Operator 3

TESTS OF INDIVIDUAL MEANS

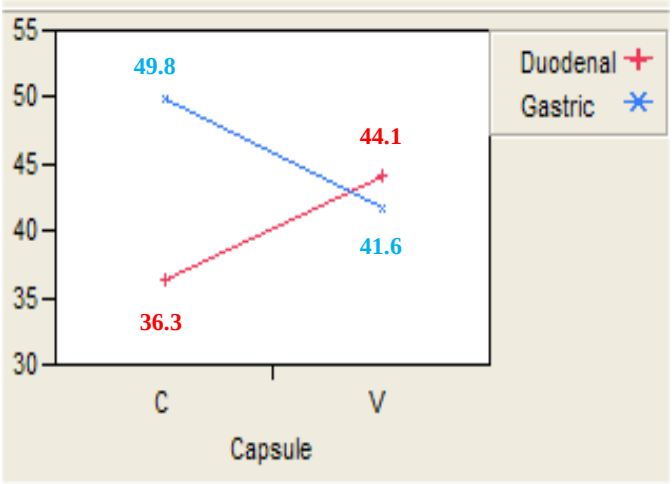
When there is no interaction, the comparisons of treatment means (either $\bar{y}_{i\bullet\bullet}$ or $\bar{y}_{\bullet j\bullet}$, depending upon whether the row-factor or the column-factor is significant) may be made using DMRT or Fisher's LSD. In such a case, for DMRT we have to use

$$S_{\bar{y}_{i\bullet\bullet}} = \sqrt{\frac{MS_E}{bn}} \text{ or } S_{\bar{y}_{\bullet j\bullet}} = \sqrt{\frac{MS_E}{an}}$$

and for LSD method the respective expressions for LSD will be

$$t_{\alpha/2,DF_E}\sqrt{\frac{2MS_E}{bn}} \text{ or } t_{\alpha/2,DF_E}\sqrt{\frac{2MS_E}{an}}.$$

Consider the following interaction plot between capsule types (C and V) and digestive fluids type (duodenal and gastric) with response as capsule dissolve time.



Question: Do you think

1. Capsule type has significant impact on response?
2. Digestive Fluid type has significant impact on response?
3. Interaction is significant?

Above shows that when interaction is significant, comparisons between the means of one factor (Capsule Type in above example) may be obscured/masked by the presence of interaction. In this case we could apply DMRT or Fisher's LSD to the means of factor Capsule Type, with factor Digestive Fluids Type set at a particular level. Here, we will have $S_{\bar{y}_{ij\cdot}} = \sqrt{\frac{MS_E}{n}}$, **for a selected** $j = j'$ **and** $LSD = t_{\alpha/2,DF_E}\sqrt{\frac{2MS_E}{n}}$.

RESIDUAL ANALYSIS

The residuals from two-way design are defined as

$$e_{ijk} = y_{ijk} - \hat{y}_{ijk} = y_{ijk} - \bar{y}_{ij\bullet}$$

$$\hat{\mu} = \bar{y}_{\cdots},$$

$$\hat{\tau}_i = \bar{y}_{i\cdot\cdot} - \bar{y}_{\cdots}, \quad \hat{\beta}_j = \bar{y}_{\cdot j\cdot} - \bar{y}_{\cdots},$$

$$(\widehat{\tau\beta})_{ij} = \bar{y}_{ij\cdot} - \bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot j\cdot} + \bar{y}_{\cdots}.$$

THREE-WAY Classification

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

$$CF = \frac{y_{\bullet\bullet\bullet\bullet}^2}{abcn}$$

$$TSS = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}\sum_{l=1}^{n} y_{ijkl}^2 - CF$$

$$SS_A = \frac{1}{bcn}\sum_{i=1}^{a} y_{i\bullet\bullet\bullet}^2 - CF$$

$$SS_B = \frac{1}{acn}\sum_{j=1}^{b} y_{\bullet j\bullet\bullet}^2 - CF$$

$$SS_C = \frac{1}{abn}\sum_{k=1}^{c} y_{\bullet\bullet k\bullet}^2 - CF$$

$$SS_{AB} = \frac{1}{cn}\sum_{i=1}^{a}\sum_{j=1}^{b} y_{ij\bullet\bullet}^2 - CF - SS_A - SS_B$$

$$SS_{AC} = \frac{1}{bn}\sum_{i=1}^{a}\sum_{k=1}^{c} y_{i\bullet k\bullet}^2 - CF - SS_A - SS_C$$

$$SS_{BC} = \frac{1}{an}\sum_{j=1}^{b}\sum_{k=1}^{c} y_{\bullet jk\bullet}^2 - CF - SS_B - SS_C$$

$$SS_{ABC} = \frac{1}{n}\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c} y_{ijk\bullet}^2 - CF - SS_A - SS_B - SS_C - SS_{AB} - SS_{AC} - SS_{BC}$$

$$SS_E = TSS - SS_A - SS_B - SS_C - SS_{AB} - SS_{AC} - SS_{BC} - SS_{ABC}$$

Assuming that all independent variables (i.e., *A, B,* and *C*) are fixed factors; the analysis of variance table can be formulated as shown below.

Source of Variation	SS	DF	MS	E(MS)	F_0
A	SS_A	$a-1$	MS_A	$\sigma^2+\frac{bcn\sum\tau_i^2}{a-1}$	$\frac{MS_A}{MS_E}$
B	SS_B	$b-1$	MS_B	$\sigma^2+\frac{acn\sum\beta_j^2}{b-1}$	$\frac{MS_B}{MS_E}$
C	SS_C	$c-1$	MS_C	$\sigma^2+\frac{abn\sum\gamma_k^2}{c-1}$	$\frac{MS_C}{MS_E}$
AB	SS_{AB}	$(a-1)(b-1)$	MS_{AB}	$\sigma^2+\frac{cn\sum\sum(\tau\beta)_{ij}^2}{(a-1)(b-1)}$	$\frac{MS_{AB}}{MS_E}$
AC	SS_{AC}	$(a-1)(c-1)$	MS_{AC}	$\sigma^2+\frac{bn\sum\sum(\tau\gamma)_{ik}^2}{(a-1)(c-1)}$	$\frac{MS_{AC}}{MS_E}$
BC	SS_{BC}	$(b-1)(c-1)$	MS_{BC}	$\sigma^2+\frac{an\sum\sum(\beta\tau)_{jk}^2}{(b-1)(c-1)}$	$\frac{MS_{BC}}{MS_E}$
ABC	SS_{ABC}	$(a-1)(b-1)(c-1)$	MS_{ABC}	$\sigma^2+\frac{n\sum\sum\sum(\tau\beta\gamma)_{ijk}^2}{(a-1)(b-1)(c-1)}$	$\frac{MS_{ABC}}{MS_E}$
$Error$	SS_E	$abc(n-1)$	MS_E	σ^2	
$Total$	SS_T	$abcn-1$			

Note: There must be at least two (2) replicates to compute the error sum of squares, otherwise some higher order interaction term(s) has(have) to be sacrificed for obtaining the error sum of squares.

Example:

An experiment was conducted to determine the effects of temperature, pressure, and stirring rate on product filtration rate. This was done in a pilot plant. The experiment was run at two levels of each factor. In addition, it was decided that two batches of raw materials should be used, where batches were treated as blocks. Eight experimental runs for each replication *were made in random order for each batch of raw materials*. It is thought that all two-factor interactions may be of interest. No interactions with batches are assumed to exist. The data appear in the following table. The filtration rate is in gallons per hour.

a) Show the complete ANOVA table.
b) What interactions appear to be significant?
c) Pool all insignificant interactions with error and write the revised ANOVA table.

Batch No.			Stirring Rate			
			Low		High	
			Pressure		Pressure	
			Low	High	Low	High
1	Temperature	Low	43	49	44	47
		High	64	68	97	102
2	Temperature	Low	49	57	51	55
		High	70	76	103	106

Solution

Factors	Level-wise Total	
	Level 1	Level 2
Block (Batch)	514	567
Stirring Rate (A)	476	605
Pressure (B)	521	560
Temperature (C)	395	686

Note: Each level-wise total is obtained by adding 8 observed values.

Interaction Table

A X B	A1	A2
B1	226	295
B2	250	310

A X C	A1	A2
C1	198	197
C2	278	408

B X C	B1	B2
C1	187	208
C2	334	352

- Grand Total = 1081
- Correction Factor = 73035.06
- RSS = 80725
- Total SS = 7689.938
- Block SS = 175.562
- Stirring Rate SS = 1040.063
- Pressure SS = 95.062
- Temperature SS = 5292.563
- SR X Pressure SS = 5.062
- SR X Temp SS = 1072.563
- Pressure X Temp SS = 0.562
- Error SS = 8.5

ANOVA

Factor	SS	DOF	MS	F	F Crit	Remark
Block	175.562	1	175.562	165.235		Significant
SR	1040.063	1	1040.063	978.882		Significant
Pressure	95.062	1	95.062	89.471		Significant
Temp	5292.563	1	5292.563	4981.235	5.318	Significant
SR X Pr	5.062	1	5.062	4.765		Insignificant
SR X Temp	1072.563	1	1072.563	1009.471		Significant
Pr X Temp	0.562	1	0.562	0.529		Insignificant
Error	8.5	8	1.062			
Total	7689.938	15				

So, interaction between Stirring Rate and Temperature appears to have significant impact on filtration rate.

So, adding other two interactions SS, namely Stirring Rate × Pressure and Pressure × Temperature, to Error SS and corresponding DF's to Error DF, we get following revised ANOVA.

Revised (Pooled) ANOVA

Factor	SS	DOF	MS	F	F Crit	Remark
Block	175.562	1	175.562	124.292		Significant
SR	1040.063	1	1040.063	736.327		Significant
Pressure	95.062	1	95.062	67.301	4.965	Significant
Temp	5292.563	1	5292.563	3746.947		Significant
SR X Temp	1072.563	1	1072.563	759.336		Significant
Pooled Error	14.125	10	1.4125			
Total	7689.938	15				

RULES FOR DETERMINATION OF EXPECTED MEAN SQUARES (EMS)

- **Models could be a) Fixed Effect Model, b) Random Effect Model, and c) Mixed Effect Model.**

- **Replicates are always considered as random.**

- **If an interaction contains at least one random effect, the entire interaction is considered to be a random effect**

- **Represent the error term ε_{ijk} by $\varepsilon_{(ij)k}$, where subscripts in brackets are called 'dead' subscript, whereas those outside the brackets are known as 'live' subscript.**

- **Degrees of freedom for the main effects of factors that contains**
  - **dead subscript: number of levels, and**
  - **live subscript: number of levels – 1.**

- **Degrees of freedom for any interaction term in the model is the product of the degrees of freedom of the corresponding main effects. For example: DF of $(\tau\beta)_{ij} = (a-1)(b-1)$ and DF of $\varepsilon_{(ij)k} = ab(n-1)$.**

- Each model term has
  - i) either a variance component (for random effect), denoted by σ_{α}^2 or σ_{β}^2 or $\sigma_{\alpha\beta}^2$,
  - ii) or a fixed factor (for fixed effect) that is represented by sum of squares of the corresponding model parameter divided by its degrees of freedom, i.e. $\frac{1}{a-1}\sum\tau_i^2$ corresponding to model parameter τ_i, $1 \le i \le a$. or $\frac{\sum_{i=1}^{a}\sum_{j=1}^{b}(\tau\beta)_{ij}^2}{(a-1)(b-1)}$ for model parameter $(\tau\beta)_{ij}$.

- Prepare a table, depending upon the model used, as below.
  Let the model be

$$y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \varepsilon_{(ij)k} \begin{cases} i = 1, 2, \cdots, a \\ j = 1, 2, \cdots, b, \\ k = 1, 2, \cdots, n \end{cases}$$

  where both the factors are assumed to be fixed.

No. of levels	a	b	n
Fixed or Random Factor	F	F	R
Factor Subscript	i	j	k
τ_i			
β_j			
$(\tau\beta)_{ij}$			
$\varepsilon_{(ij)k}$			

- In each cell, write '1' if the dead subscript in the row component matches the subscript in the column.

- In each cell, if any of the live subscripts of the row component match the subscript in the column, write '0' if the column is headed by a fixed factor and '1' if the column is headed by a random factor.

- Fill the remaining empty cells with corresponding number of levels shown in the column heading.

- So, we get the following table:

No. of levels	a	b	n
Fixed or Random Factor	F	F	R
Factor Subscript	i	j	k
τ_i	0	b	n
β_j	a	0	n
$(\tau\beta)_{ij}$	0	0	n
$\varepsilon_{(ij)k}$	1	1	1

- To obtain the expected mean square for any "model term", first cover all columns headed by live subscripts associated with that term. Then, consider those rows that contains at least the same subscripts as that of the "model term", take the product of the visible numbers per row and multiply the resulting product by the appropriate fixed factor or variance component (as discussed earlier) and add them. The resulting expression is the EMS of the "model term".

To find E(MSA), for example, cover column 1(*i* is the index associated with factor A). The product of the visible numbers in the rows that contain at least subscript *i* are bn (row 1), $n\times 0$ (row 3), and 1 (row 4). Note that *i* is missing in row 2. So, the expected mean square is

$$E\left(MS_A\right)=\sigma^2+bn\frac{\sum_{i=1}^{a}\tau_i^2}{a-1}.$$

- **The complete table for EMS for the model selected is**

No. of levels Fixed or Random Factor Factor Subscript	a F i	b F j	n R k	Expected Mean Square
τ_i	0	b	n	$E\left(MS_A\right)=\sigma^2+bn\frac{\sum_{i=1}^{a}\tau_i^2}{a-1}$
β_j	a	0	n	$E\left(MS_B\right)=\sigma^2+an\frac{\sum_{j=1}^{b}\beta_j^2}{b-1}$
$(\tau\beta)_{ij}$	0	0	n	$E\left(MS_{AB}\right)=\sigma^2+n\frac{\sum_{i=1}^{a}\sum_{j=1}^{b}(\tau\beta)_{ij}^2}{(a-1)(b-1)}$
$\varepsilon_{(ij)k}$	1	1	1	σ^2

So, testing for all the effects, i.e. all Mean Squares, are to be done w.r.t. Mean Square Error (MS_E).

EMS for Two Factor Random Effect Model

No. of levels Fixed or Random Factor Factor Subscript	a R i	b R j	n R k	Expected Mean Square
τ_i	1	b	n	$E(MS_A) = \sigma^2 + n\sigma^2_{\tau\beta} + bn\sigma^2_{\tau}$
β_j	a	1	n	$E(MS_B) = \sigma^2 + n\sigma^2_{\tau\beta} + an\sigma^2_{\beta}$
$(\tau\beta)_{ij}$	1	1	n	$E(MS_{AB}) = \sigma^2 + n\sigma^2_{\tau\beta}$
$\varepsilon_{(ij)k}$	1	1	1	$E(MS_E) = \sigma^2$

So, here "interaction" effect is to be tested against MS_E. Now, for testing main effects there could be two situations:

i) Interaction is Significant: Test with $MS_{\text{Interaction}}$, Or

ii) Interaction is Insignificant: Test with pooled estimate of MS_E, obtained by dividing $(SS_{Interaction} + SS_E)$ by $(DF_{Interaction} + DF_E)$.

➢ **Find the Expected Mean Squares for Two Factors Mixed Model, where Factor A is fixed, and Factor B is random and derive the testing procedure.**

No. of levels Fixed or Random Factor Subscript	a F i	b R j	n R k	Expected Mean Square
τ_i	0	b	n	$E(MS_A) = \sigma^2 + n\sigma^2_{\tau\beta} + bn\sum\tau_i^2/(a-1)$
β_j	a	1	n	$E(MS_B) = \sigma^2 + an\sigma^2_{\beta}$
$(\tau\beta)_{ij}$	0	1	n	$E(MS_{AB}) = \sigma^2 + n\sigma^2_{\tau\beta}$
$\varepsilon_{(ij)k}$	1	1	1	$E(MS_E) = \sigma^2$

Approximate F Tests

In factorial experiments with three or more factors, where factors are random or mixed, there are frequently no exact test statistics for certain effects in the model.

Consider three-factor factorial experiment with a levels of factor A, b levels of factor B, c levels of factor C, and n replicates where all the factors are *random*. The appropriate statistical model is:

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

Expected mean squares for above model is given below.

Model Term	Factor	Expected Mean Square
τ_i	**A,** main effect	$\sigma^2 + cn\sigma^2_{\tau\beta} + bn\sigma^2_{\tau\gamma} + n\sigma^2_{\tau\beta\gamma} + bcn\sigma^2_{\tau}$
β_j	**B,** main effect	$\sigma^2 + cn\sigma^2_{\tau\beta} + an\sigma^2_{\beta\gamma} + n\sigma^2_{\tau\beta\gamma} + acn\sigma^2_{\beta}$
γ_k	**C,** main effect	$\sigma^2 + bn\sigma^2_{\tau\gamma} + an\sigma^2_{\beta\gamma} + n\sigma^2_{\tau\beta\gamma} + abn\sigma^2_{\tau}$
$(\tau\beta)_{ij}$	**AB,** interaction	$\sigma^2 + n\sigma^2_{\tau\beta\gamma} + cn\sigma^2_{\tau\beta}$
$(\tau\gamma)_{ik}$	**AC,** interaction	$\sigma^2 + n\sigma^2_{\tau\beta\gamma} + bn\sigma^2_{\tau\gamma}$
$(\beta\gamma)_{jk}$	**BC,** interaction	$\sigma^2 + n\sigma^2_{\tau\beta\gamma} + an\sigma^2_{\beta\gamma}$
$(\tau\beta\gamma)_{ijk}$	**ABC,** interaction	$\sigma^2 + n\sigma^2_{\tau\beta\gamma}$
ε_{ijkl}	**Error**	σ^2

We notice by examining the expected mean squares in above table, that if *A*, *B*, and *C* are all random factors, then no exact test exists for the main effects.

In such cases, a procedure attributed to Satterthwaite (1946) can be employed. Satterthwaite's method uses linear combination of mean squares, for example

$$MS' = MS_r + \cdots + MS_s$$

and

$$MS'' = MS_u + \cdots + MS_v$$

where the mean square in above equations are so chosen that $E(MS') - E(MS'')$ becomes equal to a multiple of the effect (the model parameter or variance component) considered in the respective null hypothesis. Then the test statistic would be

$$F = \frac{MS'}{MS''}$$

which is distributed, approximately, as $F_{p,q}$ where

$$p = \frac{(MS_r + \cdots + MS_s)^2}{MS_r^2/df_r + \cdots + MS_s^2/df_s}$$

and

$$q = \frac{(MS_u + \cdots + MS_v)^2}{MS_u^2/df_u + \cdots + MS_v^2/df_v}$$

For example, in the three factor random effects model, an appropriate test statistic for H_0: $\sigma_\tau^2 = 0$ would be $F = MS'/MS''$, with

$$MS' = MS_A + MS_{ABC}$$

and

$$MS'' = MS_{AB} + MS_{AC}.$$

This is because

$$E(MS') = E(MS_A + MS_{ABC}) = 2\sigma^2 + 2n\sigma^2_{\tau\beta\gamma} + cn\sigma^2_{\tau\beta} + bn\sigma^2_{\tau\gamma} + bcn\sigma^2_{\tau}$$

and

$$E(MS'') = E(MS_{AB} + MS_{AC}) = 2\sigma^2 + 2n\sigma^2_{\tau\beta\gamma} + cn\sigma^2_{\tau\beta} + bn\sigma^2_{\tau\gamma}$$

so that $E(MS') - E(MS'') = bcn\sigma^2_{\tau}$.

The degrees of freedom for *F* would be computed from expressions given earlier.

Similarly, we can derive appropriate approximate F test statistic for testing main effects of factors B and C.

Example>

Consider the three-factor factorial model

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

Assuming all the factors A, B, and C are random, develop the ANOVA Table, including the expected mean squares. Propose appropriate test statistics for all effects.

Solution>

Formulae for sum of squares:

Sum of Squares	Expression for SS
TSS	$\sum\sum\sum y_{ijk}^2 - \frac{y_{...}^2}{abc}$
SS(A)	$\frac{1}{bc}\sum y_{i..}^2 - \frac{y_{...}^2}{abc}$
SS(B)	$\frac{1}{ac}\sum y_{.j.}^2 - \frac{y_{...}^2}{abc}$
SS(C)	$\frac{1}{ab}\sum y_{..k}^2 - \frac{y_{...}^2}{abc}$
SS(AB)	$\frac{1}{c}\sum y_{ij.}^2 - \frac{y_{...}^2}{abc} - SS(A) - SS(B)$
SS(BC)	$\frac{1}{a}\sum y_{.jk}^2 - \frac{y_{...}^2}{abc} - SS(B) - SS(C)$
SS(Error)	By Subtraction

ANOVA table with mean sum of squares:

Factor	DF	SS	E(MS)
A	a -1	SS(A)	$\sigma^2 + c\sigma^2_{\tau\beta} + bc\sigma^2_{\tau}$
B	b -1	SS(B)	$\sigma^2 + a\sigma^2_{\beta\gamma} + c\sigma^2_{\tau\beta} + ac\sigma^2_{\beta}$
C	c -1	SS(C)	$\sigma^2 + a\sigma^2_{\beta\gamma} + ab\sigma^2_{\gamma}$
AB	(a -1)(b -1)	SS(AB)	$\sigma^2 + c\sigma^2_{\tau\beta}$
BC	(b-1)(c-1)	SS(BC)	$\sigma^2 + a\sigma^2_{\beta\gamma}$
Error	b(a-1)(c-1)	SS(*Error*)	σ^2
Total	abc -1		

So, exact tests exist for all effects except the main effect B.

To test the main effect B, or H_0: $\sigma^2_{\beta} = 0$, we could use the statistic

$$\boldsymbol{F = \frac{MS'}{MS''}}$$

where

$\boldsymbol{MS' = MS_B + MS_E}$ **and** $\boldsymbol{MS'' = MS_{AB} + MS_{BC}}$ **.**

ANOVA and Measurement System Analysis

Measurement System Analysis is an experimental and mathematical method of determining how much the variation within the measurement process contributes to overall process variability. It may be noted that experiments in such a case should be a completely randomized one.

Two major components of MSA are: repeatability and reproducibility.

Repeatability is the extent of variation when same appraiser (operator) measures same part repeatedly using same instrument/gauge.

Reproducibility is the extent of variation when different appraisers/ operators measure same part repeatedly using same instrument/gauge.

Measurement System Metrics

Measurement Error is sum of the variations due to repeatability and reproducibility, i.e.

$$\sigma_{measurement\ error} = \sqrt{\sigma^2_{repeatability} + \sigma^2_{reproducibility}}$$

Total Measurement Error is the interval that contains 99% of the probable measurement values, using single part, from a measurement system, i.e.

total measurement error = $5.15 \times \sigma_{measurement\ error}$

Measurement System Precision is defined Precision-to-Tolerance Ratio (P/T Ratio), which can be expressed as

$$P/T\ Ratio\ (\%) = \frac{5.15 \times \sigma_{measurement\ error}}{USL - LSL} \times 100$$

According to Automotive Industry Action Group, AIAG (2002), a general rule of thumb for measurement system acceptability in terms of P/T ratio (%), is:

Excellent	Acceptable	Unacceptable
< 10%	10 – 30 %	> 30%

Example> A company produces powdered resin that includes a silica modifier to improve dimensional stability. Specification of silica concentration by weight is 15 ± 2%. A process control engineer randomly collected five powder samples, which span the silica specification range and included two out-of-specification samples pulled out from quarantine lots. One composite sample is made from each powder bag. Randomly selected three appraisers are told to analyse same composite sample, made from each powder bag, for silica content thrice each. The order in which measurements are made is completely randomized. The following table gives the data thus obtained.

Sample Bag #	Appraiser 1			Appraiser 2			Appraiser 3		
	Trial 1	Trial 2	Trial 3	Trial 1	Trial 2	Trial 3	Trial 1	Trial 2	Trial 3
1	18.2	17.9	18.2	18.1	18.0	18.0	17.8	17.8	18.2
2	14.4	14.9	14.8	14.8	14.6	14.8	14.4	14.4	14.5
3	14.0	13.9	13.8	13.9	14.2	14.0	13.8	13.7	13.8
4	17.2	17.2	17.4	17.4	17.3	17.5	17.4	17.5	17.5
5	12.9	12.8	12.5	12.5	12.9	12.7	12.9	12.5	12.6

Above is clearly a case two-factor factorial experiment with three replications where both the design factors samples (5) and appraisers (3) are random factors. So, here $a = 5, b = 3, \text{and } n = 3.$

The statistical model is: $y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \epsilon_{ijk}$, **where** i, j, k **respectively represents sample bag, appraiser and replication.**

So, the corresponding variance component identity is

$$\sigma_y^2 = \sigma_\tau^2 + \sigma_\beta^2 + \sigma_{\tau\beta}^2 + \sigma^2$$

Typically,

$$\sigma^2_{repeatability} = \sigma^2, \text{ and}$$

$$\sigma^2_{reproducibility} = \begin{cases} \sigma_\beta^2 + \sigma_{\tau\beta}^2 \\ \sigma_\beta^2, \quad \text{if interaction is not significant} \end{cases}$$

Gauge R&R Study – ANOVA Method

Please note here that (refer two-factor random effect model)

$$E(MS_{SB}) = \sigma^2 + 3\sigma_{\tau\beta}^2 + (3 \times 3)\sigma_\tau^2$$
$$E(MS_{App}) = \sigma^2 + 3\sigma_{\tau\beta}^2 + (5 \times 3)\sigma_\beta^2$$
$$E(MS_{SB\times App}) = \sigma^2 + 3\sigma_{\tau\beta}^2$$
$$E(MS_{Error}) = \sigma^2$$

Two-way ANOVA Table with Interaction

Source	DF	SS	MS	F	F -Critical
Sample Bag	4	188.1364	47.0341	1479.1	3.84
Appraiser	2	0.1258	0.0629	1.978	4.46
Interaction	8	0.2542	0.0318	1.314	2.27
Error	30	0.7267	0.0242		
Total	44	189.2431			

Note: $F_{Int} = \frac{MS_{Int}}{MS_E}$, $F_{Appraiser} = \frac{MS_{App}}{MS_{Int}}$, $F_{SB} = \frac{MS_{SB}}{MS_{Int}}$.

Two-way ANOVA Table without Interaction

Source	DF	SS	MS	F	F - Critical
Sample Bag	4	188.1364	47.0341	1823.027	2.626
Appraiser	2	0.1258	0.0629	2.438	3.248
Error (Pooled)	38	0.9809	0.0258		
Total	44	189.2431			

$$\sigma^2_{repeatability} = \hat{\sigma}^2 = 0.0258$$

Now, since interaction is insignificant, we will have

$$E\left(MS_{App}\right) = \sigma^2 + (5 \times 3)\sigma^2_{\beta}.$$

So,

$$\sigma^2_{reproducibility} = \frac{0.0629 - 0.0258}{5 \times 3} = 0.00247$$

Gauge R&R Summary

Source	Std. Deviation	$5.15 \times SD$	P/T Ratio (%)
Repeatability	0.1606	0.8272	20.68
Reproducibility	0.0497	0.2560	6.40
Gauge R&R	0.1681	0.8657	21.64

Note: USL – LSL = 17 - 13 = 4.

Exercise 1

An experiment was run to determine whether four specific firing temperatures affect the density of a certain type of brick. The experiment led to the following data.

Temperature $\left({}^{0}F\right)$	Density						
100	21.8	21.9	21.7	21.6	21.7	21.5	21.8
125	21.7	21.4	21.5	21.5			
150	21.9	21.8	21.8	21.6	21.5		
175	21.9	21.7	21.8	21.7	21.6	21.8	

a) Does the firing temperature affect the density of the bricks? Use $\alpha = 0.05$.
b) Find the 95% confidence interval for the mean density at 125 ^{0}F.

Exercise 2

The response time in milliseconds was determined for three different types of circuits in an electronic calculator. The results are recorded below.

Circuit Type	Response				
1	19	22	20	18	25
2	30	31	33	27	40
3	16	15	18	26	17

a) Using $\alpha = 0.01$, test the hypothesis that the three circuit types have the same response time.
b) Use Fisher's LSD method to analyze the mean response time for the three circuits. Use $\alpha = 0.01$.

Exercise 3

A textile manufacturing company weaves a fabric on a large number of looms. The company is interested in the loom-to-loom variability in tensile strength of the fabric. To investigate this, a manufacturing engineer selects four looms at random and makes four strength determinations on fabric sample chosen at random from each loom. The data are shown below.

Loom	Tensile strength values			
	1	2	3	4
1	98	97	99	96
2	91	90	93	92
3	96	95	97	95
4	95	96	99	98

a) Are the looms being similar in their ability to produce fabric of equal strength?
b) Estimate the variance components, that is, variability due to treatment effects and residual/error variation.

Exercise 4

An experiment is conducted to investigate warping of copper plates. The two factors studied were temperature and the copper content of the plates. The response variable is the amount of warping. The data obtained on amount of warping are as follows.

Temperature (^{0}C)	Copper Content (%)			
	40	60	80	100
50	17, 20	16, 21	24, 22	28, 27
75	12, 9	18, 13	17, 12	27, 31
100	16, 12	18, 21	25, 23	30, 23
125	21, 17	23, 21	23, 22	29, 31

a) Is there any indication that either factor affects the amount of warping? Is there any interaction between the factors?
b) Compare the average warping at each level of copper content using Fisher's LSD method.
c) If low warping is desirable, what level of the factors would you suggest?

Exercise 5

A mechanical engineer is studying the surface roughness of a part produced in a metal-cutting operation. Three factors – feed rate (*A*), depth of cut (*B*) and tool angle (*C*) are of interest. All three factors have been assigned two levels and two replicates of a factorial design are run. The coded data are shown in the table below.

Feed Rate (A)	Depth of Cut (B)			
	0.025 inch		0.040 inch	
	Tool Angle (C)		Tool Angle (C)	
	15^0	25^0	15^0	25^0
20 inches /min	9, 7	11, 10	9, 11	10, 8
30 inches /min	10,12	10, 13	12, 15	16, 14

Analyze the data using analysis of variance assuming that all factors are fixed. Use $\alpha = 0.05$.

Exercise 6.

The percentage of hardwood concentration in raw pulp, the vat pressure, and the cooking time of pulp are being investigated for their effects on the strength of paper. Three levels of hardwood concentration, three levels of pressure and two cooking times are selected. A factorial experiment with two replicates is conducted and the following data are obtained.

% Hardwood Concentration	Cooking Time 3 hrs. Pressure			Cooking Time 4 hrs. Pressure		
	400	500	650	400	500	650
2	196.6	197.7	199.8	198.4	199.6	200.6
	196	196	199.4	198.6	200.4	200.9
4	198.5	196	198.4	197.5	198.7	199.6
	197.2	196.9	197.6	198.1	198.1	199
8	197.5	195.6	197.4	197.6	197.6	198.5
	196.6	196.2	198.1	198.4	198.4	199.8

a) Analyze the data assuming hardwood concentration and cooking times are fixed while pressure is random.
b) Analyze the same data assuming that all the factors are random.

Exercise 7

An organization manufacturing mobile phones is interested in finding amount variability that exist in the measurement process in one of its critical parameters – luminance of the cell phone screen. With that in view an experiment was conducted with 8 randomly selected cell phones of a specific make and 3 randomly chosen operators. Each operator was told to measure luminance of each cell phone, using photometer, twice. Order of trial for collection of luminance data was completely randomized. Following table gives the data on luminance:

Part # (cell phone)	Operator 1		Operator 2		Operator 3	
	Trial 1	Trial 2	Trial 1	Trial 2	Trial 1	Trial 2
1	0.65	0.6	0.55	0.55	0.5	0.55
2	1	1	1.05	0.95	1.12	1
3	0.85	0.8	0.8	0.75	0.8	0.8
4	0.85	0.95	0.8	0.75	0.8	0.8
5	0.55	0.45	0.4	0.4	0.45	0.5
6	1	1	1	1.05	1	1.05
7	0.95	0.95	0.95	0.9	0.95	0.95
8	0.85	0.8	0.75	0.7	0.8	0.8

The specification for luminance is 0.5 to 1.5. Analyze the data suitably. Calculate the P/T ratio and comment on the acceptability of the measurement system.