ANALYSIS OF VARIANCE

(A case study demonstrating the applications of one-way, two-way and incomplete three-way ANOVA, one-way and two-way ANOCOVA, and Factorial Experiments for some industrial and agricultural data.)

Submitted by: Vishal Kumar

Due Date: 31st January, 2016

Date of Submission: 31st January, 2016

Instructor's Remarks:

4. Analysis of Variance (10 Cases)

4.1 One-Way ANOVA

4.1.1 Completely Randomized Design – Analyzing performance of 4 different Drugs over Placebo

4.2 Two-Way ANOVA

4.2.1 Randomized Block Design (RBD) with 1 observation per cell – Analyzing three different Methods of Analysis while controlling for the effect of Analysts

4.2.2 Randomized Block Design (RBD) with m observations per cell – Analyzing the effect of Variety, Spacing Pattern, and their interaction on the crop yield

4.2.3 Latin Square Design (LSD) – Comparing the effects of 5 different Chemicals while controlling for the effect of Day and Batch of experiment

4.3 One-Way ANCOVA – Comparing Breaking Strength of Fiber produced by three different Machines while controlling for the effect of Diameter of the Fiber

4.4 Two-Way ANCOVA – Comparing the effect of five Manure Treatments on Cotton Crop Yield while controlling for Fertility and No. of Plants per plot

4.5 Factorial Experiments

4.5.1 2^3 Factorial Experiment – analyzing the effect of Nitrogen, Phosphorus and Potassium and their all possible interactions on Potato Crop Yield

4.5.2 2^3 Factorial Experiment with Total Confounding - analyzing the effect of Nitrogen, Phosphorus and Potassium and their interactions on Potato Crop Yield while confounding the highest order interaction

4.5.3 2^3 Factorial Experiment with Partial Confounding - analyzing the effect of Nitrogen, Phosphorus and Potassium and their all possible interactions on Potato Crop Yield while confounding different interactions in different Replicates

4.5.4 3^2 Factorial Experiment – Analyzing the effect of Developer Strength and Development Time on the density of Photographic Film

Analysis of Variance

Analysis of variance (ANOVA) is a collection of statistical models used to analyse the differences among group means and their associated procedures (such as "variation" among and between groups), developed by statistician Ronald A. Fisher. In the ANOVA, the observed variance in a particular variable is partitioned into components attributable to different sources of variation. In its simplest form, ANOVA provides a statistical test of whether or not the means of several groups are equal, and therefore generalizes the t-test to more than two groups. ANOVAs are useful for comparing (testing) three or more means (groups or variables) for statistical significance. It is conceptually similar to multiple two-sample t-tests, but is less conservative (results in less type I error) and is therefore suited to a wide range of practical problems.

Assumptions in ANOVA:

The analysis of variance can be presented in terms of a linear model, which makes the following assumptions:

- Independence of observations – this is an assumption of the model that simplifies the statistical analysis.
- Normality – the distributions of the residuals are normal.
- Equality (or "homogeneity") of variances, called homoscedasticity — the variance of data in groups should be the same.

The separate assumptions of the model imply that the errors are independently, identically, and normally distributed for fixed effects models.

In this case study we have covered Completely Randomized Design, Randomized Block Design with one and more than one observations per cell, Latin square design and various Factorials experiments with two and three level of factors and perform One way ANOVA, Two ways ANOVA, One way ANACOVA and Two way ANACOVA.

One-Way ANOVA

Completely Randomized Design

Case 1: Suppose that in an investigation of the effects of 4 supposedly performance enhancing drugs upon the skilled performance is tested on 5 groups of subjects. Does any one of the drugs affect the performance levels?

Treatment	Performance Levels									
Placebo	10	9	7	9	11	5	7	6	8	8
Drug A	8	10	7	7	7	12	7	4	9	8
Drug B	12	14	9	7	15	12	14	14	11	12
Drug C	13	12	17	12	10	24	13	13	20	12
Drug D	11	20	15	6	11	12	15	15	12	12

In design of experiment terms we have we have 5 treatments. We want to performance enhancing drugs affect the performance level. We have 10 subjects in each group which makes in total 50 observation.

We want to compare 5 treatments, hence we will use Completely Randomized Design. This is the most simplest and flexible design. In this design the experimental units are allotted at random to the treatments, so that every unit get the same chance of receiving every treatment.

$$\boldsymbol{H_o}: m_P = m_A = m_B = m_C = m_D \quad \text{i.e. Homogeneity between treatment effects.}$$

$$\boldsymbol{H_a}: at\ least\ one\ of\ treatment\ mean\ differ\ significantly.$$

Where m_p= treatment mean of placebo

m_A = treatment mean of drug A

m_B = treatment mean of drug B

m_C = treatment mean of drug C

m_D = treatment mean of drug D

The model can be written in the form

$$y_{ij} = \mu + \tau_i + €_{ij}, \qquad i = 1, 2, \dots t, ;\ j = 1, 2, \dots N$$

where the terms of the model are defined as follows:

y_{ij} : Observation on j^{th} experimental unit receiving treatment i.

μ : Overall treatment mean, an unknown constant.

τ_i : An effect due to treatment i.

$€_{ij}$: A random error associated with the response from the j^{th} experimental unit receiving treatment i. We require that the $€_{ij}$ have a normal distribution with mean 0 and common variance σ^2. In addition, the errors must be independent.

So performing one way ANOVA of CRD give the result:

ANOVA

Observation

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	361.480	4	90.370	9.678	.000
Within Groups	420.200	45	9.338		
Total	781.680	49			

Result: We observe p value is < 0.05, hence we reject H_0 at 5% level of significance and conclude **treatment** effects differ significantly.

We now proceed to Post-Hoc Analysis to find out which pair of treatments differ significantly.

Multiple Comparisons

Dependent Variable: Observation

Tukey HSD

(I) Treatment	(J) Treatment	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
1	2	.10000	1.36659	1.000	-3.7831	3.9831
	3	-4.00000*	1.36659	.041	-7.8831	-.1169
	4	-6.60000*	1.36659	.000	-10.4831	-2.7169
	5	-4.90000*	1.36659	.007	-8.7831	-1.0169
2	1	-.10000	1.36659	1.000	-3.9831	3.7831
	3	-4.10000*	1.36659	.034	-7.9831	-.2169
	4	-6.70000*	1.36659	.000	-10.5831	-2.8169
	5	-5.00000*	1.36659	.006	-8.8831	-1.1169
3	1	4.00000*	1.36659	.041	.1169	7.8831
	2	4.10000*	1.36659	.034	.2169	7.9831
	4	-2.60000	1.36659	.331	-6.4831	1.2831
	5	-.90000	1.36659	.964	-4.7831	2.9831
4	1	6.60000*	1.36659	.000	2.7169	10.4831
	2	6.70000*	1.36659	.000	2.8169	10.5831
	3	2.60000	1.36659	.331	-1.2831	6.4831
	5	1.70000	1.36659	.726	-2.1831	5.5831
5	1	4.90000*	1.36659	.007	1.0169	8.7831
	2	5.00000*	1.36659	.006	1.1169	8.8831
	3	.90000	1.36659	.964	-2.9831	4.7831
	4	-1.70000	1.36659	.726	-5.5831	2.1831

*. The mean difference is significant at the 0.05 level.

Result: We observe pair of **treatments** (1, 3), (1,4) ,(1, 5), (2,3), (2,4), (2,5) differ significantly.

Two-Way ANOVA

Randomized Block Design with One observation per cell

Case 2: Three different methods of analysis M_1, M_2, M_3 are used to determine in parts per million the amount of a certain constituent in the sample. Each method is used by five analysts and the results are given below:

Method of Analysis			
Analyst	M_1	M_2	M_3
1	7.5	7.0	7.1
2	7.4	7.2	6.7
3	7.3	7.0	6.9
4	7.6	7.2	6.8
5	7.4	7.1	6.9

A. Do these results indicate a significant variation either between the methods or between the analysts?
B. Recode the data by changing the origin and scale as follows: u_{ij} = 10(yij − 6.5).
C. Perform ANOVA on the transformed data and compare the results with those obtained in part (a) for the original data. If there is a significant difference between the levels of any of the factors of variation, then determine the corresponding pairs of sample means which differ significantly.

Here we have 5 analyst and 3 methods of analysis. Here we have 2 sources of variations hence we use Randomized Block Design.

A randomized blocks design that treatments applied randomly in each block. In the most common situation each treatment appears once in each block. Assume there are r blocks and t treatments and there will be one observation per experimental unit. Because each of the t treatments is applied in each of the r blocks, there are tr experimental units altogether.

Letting y_{ij} denote the response from the experimental unit that received treatment i in block j, the equation for the model is

$$y_{ij} = \mu + t_i + b_j + e_{ij} \text{ where } \quad i=1,...,t;\ j=1,...,r$$

μ and t_i are fixed parameters

b_i is the random effect associated with the j^{th} block

e_{ij} is random error associated with the experimental unit in block j that received treatment i.

$\boldsymbol{H_o}: m_1 = m_2 = m_3$ (all treatment means are same i.e. homogeneity between the treatment effects)

$$\boldsymbol{H_a}: at\ least\ one\ of\ treatment\ mean\ is\ different.$$

Where m_1= treatment mean of method 1

m_2 = treatment mean of method 2

m_3 = treatment mean of method 3.

So performing one way ANOVA of RBD give the result:

Tests of Between-Subjects Effects

Dependent Variable: Observation

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	.839[a]	6	.140	8.142	.005
Intercept	764.694	1	764.694	44545.282	.000
Analyst	.043	4	.011	.621	.660
Method	.796	2	.398	23.184	.000
Error	.137	8	.017		
Total	765.670	15			
Corrected Total	.976	14			

a. R Squared = .859 (Adjusted R Squared = .754)

Result: We observe that various **analyst** do not differ significantly in determining parts per million in a given sample. We also observe that **method** used in determining parts per million in a given sample differ significantly.

We now proceed to Post-Hoc Analysis to check which pair of method differ significantly.

Multiple Comparisons

Dependent Variable: Observation

Tukey HSD

(I) Method	(J) Method	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
1	2	.3400*	.08287	.009	.1032	.5768
	3	.5600*	.08287	.000	.3232	.7968
2	1	-.3400*	.08287	.009	-.5768	-.1032
	3	.2200	.08287	.067	-.0168	.4568
3	1	-.5600*	.08287	.000	-.7968	-.3232
	2	-.2200	.08287	.067	-.4568	.0168

Based on observed means.

The error term is Mean Square(Error) = .017.

*. The mean difference is significant at the 0.05 level.

Result: We see that **method** (1, 2), (1, 3) differ significantly in determining parts per million in a given sample.

Recoding the data by changing the origin and scale as follows: $u_{ij} = 10(yij - 6.5)$ and performing ANOVA we get:

Tests of Between-Subjects Effects

Dependent Variable: Tranf_Observation

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	83.867[a]	6	13.978	8.142	.005
Intercept	614.400	1	614.400	357.903	.000
Method	79.600	2	39.800	23.184	.000
Analyst	4.267	4	1.067	.621	.660
Error	13.733	8	1.717		
Total	712.000	15			
Corrected Total	97.600	14			

a. R Squared = .859 (Adjusted R Squared = .754)

Result: We observe that various **method** do not differ significantly while various **analyst** differ significantly in determining part per million in the sample.

We now proceed to Post-Hoc Analysis to check which pair of method differ significantly.

Multiple Comparisons

Dependent Variable: Tranf_Observation

Tukey HSD

(I) Method	(J) Method	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
1	2	3.4000*	.82865	.009	1.0322	5.7678
	3	5.6000*	.82865	.000	3.2322	7.9678
2	1	-3.4000*	.82865	.009	-5.7678	-1.0322
	3	2.2000	.82865	.067	-.1678	4.5678
3	1	-5.6000*	.82865	.000	-7.9678	-3.2322
	2	-2.2000	.82865	.067	-4.5678	.1678

Based on observed means.

The error term is Mean Square(Error) = 1.717.

*. The mean difference is significant at the 0.05 level.

Result: We see that **method** (1, 2), (1, 3) differ significantly in determining parts per million in the sample.

By comparing the outputs in above 2 analysis, we see that there is no difference in the results.

Hence we conclude there is no effect of change of origin and scale.

Randomized Block Design (RBD) with m observations per cell

Case 3: An experiment has been conducted to determine the effect of 5 varieties of grains V1, V2, V3, V4 & V5 & 3 different spacing patterns S1, S2 & S3 which represent 10cms, 20cms & 30cms respectively between 2 adjacent saplings. The data below represents yield from the plots corresponding to variety-spacing combination. Test whether:

1. Different saplings are equally effective.
2. Different varieties are equally effective.
3. Different varieties of grains behave differently at different spacing.

Spacing / Variety	S1				S2				S3			
V1	56	45	43	46	60	50	45	48	66	57	50	50
V2	61	58	55	56	60	59	54	54	59	55	51	52
V3	63	53	49	48	65	56	50	50	66	58	52	55
V4	65	61	60	63	60	58	56	60	53	53	48	55
V5	60	61	50	53	62	68	67	60	73	77	77	65

In randomized block design each treatment is appear once in each block. Assume there are r blocks and t treatments, if there are m observations per cell, then we will have in all r x t x m observations.

Letting y_{ij} denote the response from the experimental unit that received treatment i in block j, the equation for the model is

$$y_{ij} = \mu + t_i + b_j + \gamma_{ij} + e_{ij} \quad \text{where} \quad i=1,\ldots,t;\ j=1,\ldots,r$$

μ and t_i are fixed parameters.

b_i is the random effect associated with the j block

γ_{ij} is the interaction effect between i^{th} treatment and j^{th} block.

$e_{\ ij}$ is random error associated with the experimental unit in block j that received treatment i.

$$\boldsymbol{H_{01}}: all\ treatment\ means\ are\ same\ \ i.e.homogeneity\ between\ the\ treatment\ effects$$

$$\boldsymbol{H_{o2}}: All\ the\ block\ effects\ are\ not\ significant.$$

$$\boldsymbol{H_{03}}: interaction\ effects\ are\ not\ significant.$$

Vs

$$\boldsymbol{H_{a1}}: at\ least\ one\ of\ treatment\ mean\ is\ different.$$

$$\boldsymbol{H_{a2}}: Block\ effects\ are\ significant.$$

$$\boldsymbol{H_{a3}}: Interaction\ effects\ are\ significant.$$

ANOVA is given by:

Tests of Between-Subjects Effects

Dependent Variable: Observation

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	2073.500[a]	14	148.107	5.694	.000
Intercept	194940.000	1	194940.000	7494.490	.000
Variety	1089.167	4	272.292	10.468	.000
Spacing	109.200	2	54.600	2.099	.134
Variety * Spacing	875.133	8	109.392	4.206	.001
Error	1170.500	45	26.011		
Total	198184.000	60			
Corrected Total	3244.000	59			

a. R Squared = .639 (Adjusted R Squared = .527)

Result: We observe **spacing** pattern between two crops does not affect the yield. **Variety** of the grain has a significant effect on the yield. But **variety** and **spacing** interact and has significant effect on the yield.

We will now proceed to Post-Hoc analysis to check which pair of variety differ significantly.

Multiple Comparisons

Dependent Variable: Observation

Tukey HSD

(I) Variety	(J) Variety	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
1	2	-4.8333	2.08211	.157	-10.7495	1.0829
	3	-4.0833	2.08211	.301	-9.9995	1.8329
	4	-6.3333*	2.08211	.030	-12.2495	-.4171
	5	-13.0833*	2.08211	.000	-18.9995	-7.1671
2	1	4.8333	2.08211	.157	-1.0829	10.7495
	3	.7500	2.08211	.996	-5.1662	6.6662
	4	-1.5000	2.08211	.951	-7.4162	4.4162
	5	-8.2500*	2.08211	.002	-14.1662	-2.3338
3	1	4.0833	2.08211	.301	-1.8329	9.9995
	2	-.7500	2.08211	.996	-6.6662	5.1662
	4	-2.2500	2.08211	.815	-8.1662	3.6662
	5	-9.0000*	2.08211	.001	-14.9162	-3.0838
4	1	6.3333*	2.08211	.030	.4171	12.2495
	2	1.5000	2.08211	.951	-4.4162	7.4162
	3	2.2500	2.08211	.815	-3.6662	8.1662
	5	-6.7500*	2.08211	.018	-12.6662	-.8338
5	1	13.0833*	2.08211	.000	7.1671	18.9995
	2	8.2500*	2.08211	.002	2.3338	14.1662
	3	9.0000*	2.08211	.001	3.0838	14.9162
	4	6.7500*	2.08211	.018	.8338	12.6662

Based on observed means.

The error term is Mean Square(Error) = 26.011.

*. The mean difference is significant at the 0.05 level.

Result: We observe **variety** (1, 4), (1, 5), (2, 5), (3, 5), (4, 5) differ significantly in determining the yield.

Latin Square Design

Case 4: Effect of 5 ingredient chemicals, Calcium (A), Nickel (B), Sulphur (C), Rubidium (D), Potassium (E), on the reaction time of a chemical process is being studied. Each batch of a new material is only large enough to permit 5 runs to be made. Furthermore, each run requires 1.5 hours so that only 5 runs can be made in one production cycle. During the day the experimenter decides to analyse the experiment as a Latin Square Design (LSD) such that the day effects & the batch effects may be systematically controlled. The experimenter obtains the following data. Analyse the data.

Day ⟶

Batch ↓

	1	2	3	4	5
1	8 (A)	7(B)	1(D)	7(C)	3(E)
2	11(C)	2(E)	7(A)	3(D)	8(B)
3	4(B)	9(A)	10(C)	1(E)	5(B)
4	6(D)	8(C)	6(E)	6(B)	10(A)
5	4(E)	2(D)	3(B)	8(A)	8(C)

In Latin square design, each treatment appears only once in each row and column , hence the variations are controlled. We have 5 chemical ingredients are expressed in the days and batch number. Hence it is a Latin square design with 5 row, columns and treatments.

$$H_o: m_A = m_B = m_C = m_D = m_E \quad \text{(Homogeneity between the treatment effects)}$$

$$H_a: at\ least\ one\ of\ treatment\ mean\ is\ different.$$

Where m_A = treatment mean of calcium

m_B = treatment mean of nickel

m_C = treatment mean of sulphur

m_D = treatment mean of rubidium

m_E = treatment mean of potassium

The model can be written in the form

$$y_{ijk} = \mu + r_i + c_j + t_k + €_{ijk}, \qquad i = 1,2, \dots m;\ j = 1,2, \dots m; k = 1,2 \dots m$$

where the terms of the model are defined as follows:

y_{ijk} : Observation in i^{th}row and j^{th} column receiving k^{th} treatment.

μ : Overall treatment mean, an unknown constant.

t_k : An effect due to treatment k, an unknown constant.

r_i: effect due to i^{th} row.

c_j: effect due to j^{th} column.

$€_{ijk}$: A random error associated with the response from the experimental unit receiving treatment k. We require that the $€_{ij}$ have a normal distribution with mean 0 and common variance σ ² . In addition, the errors must be independent.

So performing one way ANOVA of CRD give the result:

Tests of Between-Subjects Effects

Dependent Variable: Observation

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	169.120[a]	12	14.093	4.507	.007
Intercept	864.360	1	864.360	276.448	.000
Batch	15.440	4	3.860	1.235	.348
Day	12.240	4	3.060	.979	.455
Chemical	141.440	4	35.360	11.309	.000
Error	37.520	12	3.127		
Total	1071.000	25			
Corrected Total	206.640	24			

a. R Squared = .818 (Adjusted R Squared = .637)

Result: We observe **chemicals** used in a chemical reaction affects the chemical process significant. While **days** of sampling and **batch** of chemicals do not affect the chemical process significantly.

We now proceed to Post-Hoc Analysis to see which pair of chemicals differ significantly.

Multiple Comparisons

Dependent Variable: Observation

Tukey HSD

(I) Chemical	(J) Chemical	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	Upper Bound
1	2	2.8000	1.11833	.154	-.7646	6.3646
	3	-.4000	1.11833	.996	-3.9646	3.1646
	4	5.0000*	1.11833	.006	1.4354	8.5646
	5	5.2000*	1.11833	.004	1.6354	8.7646
2	1	-2.8000	1.11833	.154	-6.3646	.7646
	3	-3.2000	1.11833	.086	-6.7646	.3646
	4	2.2000	1.11833	.337	-1.3646	5.7646
	5	2.4000	1.11833	.263	-1.1646	5.9646
3	1	.4000	1.11833	.996	-3.1646	3.9646
	2	3.2000	1.11833	.086	-.3646	6.7646
	4	5.4000*	1.11833	.003	1.8354	8.9646
	5	5.6000*	1.11833	.002	2.0354	9.1646
4	1	-5.0000*	1.11833	.006	-8.5646	-1.4354
	2	-2.2000	1.11833	.337	-5.7646	1.3646
	3	-5.4000*	1.11833	.003	-8.9646	-1.8354
	5	.2000	1.11833	1.000	-3.3646	3.7646
5	1	-5.2000*	1.11833	.004	-8.7646	-1.6354
	2	-2.4000	1.11833	.263	-5.9646	1.1646
	3	-5.6000*	1.11833	.002	-9.1646	-2.0354
	4	-.2000	1.11833	1.000	-3.7646	3.3646

Based on observed means.

The error term is Mean Square(Error) = 3.127.

*. The mean difference is significant at the 0.05 level.

Result: We observe **chemicals** (1, 4), (1, 5), (3, 4), (3, 5) differ significantly and affect chemical process significantly.

One-Way ANCOVA

Case 5: 3 different machines produce fibre for a textile company. Textile engineer is interested in determining if there is a difference in breaking strength of fibre produced by 3 different machines. Assume that strength is related to diameter of the fibre. Analyse the situation and comment.

MACHINE I		MACHINE II		MACHINE III	
Y	X	Y	X	Y	X
36	20	40	22	35	21
41	24	47	27	37	23
39	25	39	22	41	27
43	25	45	30	34	27
48	32	44	29	33	16

Where Y denotes Strength and X denotes Diameter.

ANOCOVA is a technique in which it is possible to control certain sources of variations by taking additional observation on each of the experimental units. Let us suppose that in an experiment, y is the response variable and x is another variable that is linearly related to y. Moreover x cannot be controlled by the experimentor but can be observed along with y.

Let us suppose that we are comparing v treatments $t_1, t_2, \dots, t_v$,so that i^{th} treatment replicated r_i times such that $n = \sum_{i=1}^{v} r_i$ is the total number of experimental units.

Our model for ANOCOVA in CRD is:

$$y_{ij} = \mu + \alpha_i + \beta\left(x_{ij} - \bar{x}_{oo}\right) + \varepsilon_{ij} \quad ; i = 1,2,\dots,r \,; j = 1,2,\dots,r_i$$

Where

μ = general mean effect.

α_i= additional effect due to i^{th} treatment.

ε_{ij} = Random error effect.

$\beta =$ Coefficient of regression of y on x.

$x_{ij} =$ Value of concomitant variable corresponding to the response variable y_{ij}.

$\boldsymbol{H_o}: \alpha_A = \alpha_B = \alpha_C$ (All treatment means are same i.e. homogeneity between the treatment effects)

$$\boldsymbol{H_a}: at\ least\ one\ of\ treatment\ mean\ is\ different.$$

Where

$\alpha_A =$ Additional effect due to machine A

$\alpha_B =$ Additional effect due to machine B

$\alpha_C =$ Additional effect due to machine C

So performing one way ANOCOVA of CRD give the result:

Tests of Between-Subjects Effects

Dependent Variable: Y

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	240.152[a]	3	80.051	14.299	.000
Intercept	183.952	1	183.952	32.858	.000
Machine	55.802	2	27.901	4.984	.029
X	105.618	1	105.618	18.866	.001
Error	61.582	11	5.598		
Total	24462.000	15			
Corrected Total	301.733	14			

a. R Squared = .796 (Adjusted R Squared = .740)

Result: We observe that **machine** has a significant effect on the breaking strength of the fibres. We also observe that breaking strength of the fibres are not linearly associated with different **diameters**.

Two-Way ANCOVA

Case 6: In an experiment on cotton with 5 manure treatments, it was observed that the no. of plants per plot is varying from plot to plot. The yield of cotton along with the no. of plants per plot is given below. Analyse the yield data.

Replicate I	1	0	4	2	3
	12.0(24)	10.5(30)	27.0(30)	16.5(28)	25.0(35)
Replicate II	3	2	0	4	1
	26.0(40)	20.0(25)	12.0(25)	26.0(22)	15.5(28)
Replicate III	2	4	3	1	0
	22.0(32)	30.0(35)	20.0(24)	20.0(35)	14.5(30)
Replicate IV	1	3	0	4	2
	19.0(26)	18.5(16)	8.5(24)	29.0(30)	25.0(35)

Here we want to compare v treatments, each treatment replicated r times so that the total number of experimental units is $n = rv$.

Assuming a linear relationship between the response variable (y) and the concomitant variable(x), the appropriate ANOCOVA model for RBD is:

$$y_{ij} = \mu + \alpha_i + \theta_j + \beta\left(x_{ij} - \bar{x}_{oo}\right) + \mathcal{E}_{ij} \quad ; i = 1,2, \ldots, v\,; j = 1,2, \ldots, r$$

Where

μ = general mean effect

α_i= additional effect due to i^{th} treatment

θ_j = additional effect due to j^{th} block

$\mathcal{E}_{ij}$ = random error effect

β = coefficient of regression of y on x.

x_{ij} = value of concomitant variable corresponding to the response variable y_{ij}.

$$\boldsymbol{H_{o1}}: \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = \alpha_5 \quad \text{(Homogeneity between the treatment effects)}$$

$$\boldsymbol{H_{02}}: b_1 = b_2 = b_3 = b_4$$

$$\boldsymbol{H_{a1}}: at\ least\ one\ of\ treatment\ mean\ is\ different.$$

$$\boldsymbol{H_{a2}}: at\ least\ one\ of\ block\ effects\ is\ different.$$

Where

α_1 = additional effect due to manure treatment 1 ; b_1 = additional effect due to block 1

α_2 = additional effect due to manure treatment 2 ; b_2 = additional effect due to block 2

α_3 = additional effect due to manure treatment 3 ; b_3 = additional effect due to block 3

α_4 = additional effect due to manure treatment 4 ; b_4 = additional effect due to block 4

α_5 = additional effect due to manure treatment 5

So performing one way ANOCOVA of RBD give the result:

Tests of Between-Subjects Effects

Dependent Variable: Yield

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	741.185[a]	8	92.648	33.019	.000
Intercept	37.995	1	37.995	13.541	.004
Treatment	551.957	4	137.989	49.178	.000
Replicates	26.911	3	8.970	3.197	.066
nos	92.635	1	92.635	33.014	.000
Error	30.865	11	2.806		
Total	8652.500	20			
Corrected Total	772.050	19			

a. R Squared = .960 (Adjusted R Squared = .931)

Result: We observe that **treatment** affect the yield of cotton significantly. We also observe that yield of cotton is not linearly associated with the **number of plants** in a plot.

2^3 Factorial Experiment

Case 7: To study the significance of three fertilizers viz. Nitrogen (N), Phosphorus (P) and Potassium (K) on potato crop yield, an experiment was conducted where each fertilizer was taken on two different levels and repeated thrice randomly. The yield under different treatment combinations (under usual notations) are given below:

Repetition 1

NPK	(1)	K	NP	P	N	NK	PK
450	101	265	373	312	106	291	391

Repetition 2

P	NK	K	NP	(1)	NPK	PK	N
324	306	272	338	106	449	407	89

Repetition 3

P	NPK	NK	(1)	N	K	PK	NP
323	471	334	87	128	279	423	324

Test for the significance of the three fertilizers and their all possible interactions.

In this case we have three factors Nitrogen(N), Phosphorus(P) and Potassium (K) at two levels 0 and 1. Hence we have **a 2^3 factorial design**.

$$\boldsymbol{H_0}: The\ treatments\ do\ not\ have\ a\ significant\ effect$$

$$\boldsymbol{H_1}: The\ treatments\ have\ a\ significant\ effect.$$

So performing the ANOVA for 2^3 factorial experiment we get:

Tests of Between-Subjects Effects

Dependent Variable: Obs

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	349171.625[a]	9	38796.847	147.905	.000
Intercept	2012025.042	1	2012025.042	7670.423	.000
N	5673.375	1	5673.375	21.629	.000
P	205535.042	1	205535.042	783.559	.000
K	124272.042	1	124272.042	473.761	.000
N * K	1053.375	1	1053.375	4.016	.065
P * K	11837.042	1	11837.042	45.126	.000
N * P	273.375	1	273.375	1.042	.325
N * P * K	7.042	1	7.042	.027	.872
Replication	520.333	2	260.167	.992	.396
Error	3672.333	14	262.310		
Total	2364869.000	24			
Corrected Total	352843.958	23			

a. R Squared = .990 (Adjusted R Squared = .983)

Result: All the main effects are significant. We also observe that interaction between (Nitrogen, Potassium), (Nitrogen, Phosphorous) and (Nitrogen, Potassium, Phosphorous) is not significant.

2^3 Factorial Experiment with Total Confounding

Case 8: To study the significance of three fertilizers viz. Nitrogen (N), Phosphorus (P) and Potassium (K) on potato crop yield, an experiment was conducted where each fertilizer was taken on two different levels and repeated 4 times randomly with two different blocks in each replicate. The yield under different treatment combinations (under usual notations) are given below:

Repetition 1

Block – I				Block – II			
(1)	NK	NP	PK	NPK	N	K	P
101	291	373	391	450	106	265	312

Repetition 2

Block – I				Block – II			
(1)	NK	NP	PK	NPK	N	K	P
106	306	338	407	449	89	272	324

Repetition 3

Block – I				Block – II			
(1)	NK	NP	PK	NPK	N	K	P
87	334	324	423	417	128	279	323

Repetition 4

Block – I				Block – II			
(1)	NK	NP	PK	NPK	N	K	P
131	272	361	245	437	103	302	324

Test for the significance of the three fertilizers and their all possible interactions. Identify the confounded effects, if there are any. Comment on the changes in analysis and interpretation.

The eight treatment combinations require eight units of homogenous material each to form a block. Her we have a block of 4 units each, then a full replication will need only two blocks. This is done to make the blocks more homogeneous. This is called confounding. When the same treatments are grouped to form a block it is known as complete confounding.

We find that the factor **NPK is confounded** in each block.

$$H_o\text{: The complete confounding is not effective.}$$

$$H_1\text{: The complete confounding is effective.}$$

The ANOVA of the various treatments is given as:

Tests of Between-Subjects Effects

Dependent Variable: Obs

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	406943.500[a]	13	31303.346	26.098	.000
Intercept	2565112.500	1	2565112.500	2138.584	.000
Block	5080.000	6	846.667	.706	.649
N	7688.000	1	7688.000	6.410	.021
P	233928.000	1	233928.000	195.030	.000
K	127512.500	1	127512.500	106.310	.000
N * K	1922.000	1	1922.000	1.602	.222
P * K	27612.500	1	27612.500	23.021	.000
N * P	2888.000	1	2888.000	2.408	.138
N * P * K	.000	0	.	.	.
Error	21590.000	18	1199.444		
Total	2993646.000	32			
Corrected Total	428533.500	31			

a. R Squared = .950 (Adjusted R Squared = .913)

Result: We observe **block** effect is not significant hence we conclude that confounding is not effective. The main effects are significant i.e**. N, P, K** and interaction **PK.**

2^3 Factorial Experiment with Partial Confounding

Case 9: To study the significance of three fertilizers viz. Nitrogen (N), Phosphorus (P) and Potassium (K) on potato crop yield, an experiment was conducted where each fertilizer was taken on two different levels and repeated 4 times randomly with two different blocks in each replicate. The yield under different treatment combinations (under usual notations) are given below:

Repetition 1

Block – I				Block – II			
(1)	PK	NK	NP	NPK	N	K	P
55	342	234	125	450	99	78	76

Repetition 2

Block – III				Block – IV			
(1)	K	NP	NPK	NK	PK	N	P
106	234	233	530	230	78	89	90

Repetition 3

Block – V				Block – VI			
(1)	N	PK	NPK	K	P	NK	NP
87	443	234	546	456	111	200	305

Repetition 4

Block – VII				Block – VIII			
(1)	P	NK	NPK	K	N	PK	NP
133	256	389	459	423	120	320	325

Test for the significance of the three fertilizers and their all possible interactions. Identify the confounded effects, if there are any. Comment on the changes in analysis and interpretation.

When the treatments which are grouped to form a block differ in each block. Hence it is known as **partial confounding.**

A close examination reveals that the factor confounded are:

NPK is confounded in repetition 1.

NP is confounded in repetition 2.

PK is confounded in repetition 3.

NK is confounded in repetition 4.

$$\boldsymbol{H_0}: The\ partial\ confounding\ is\ not\ effective.$$

$$\boldsymbol{H_1}: The\ partial\ confounding\ is\ effective.$$

The ANOVA of the various treatments is given as:

Tests of Between-Subjects Effects

Dependent Variable: Obs

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	552546.500[a]	14	39467.607	4.040	.004
Intercept	1928648.000	1	1928648.000	197.422	.000
Block	134798.000	7	19256.857	1.971	.120
N	90100.125	1	90100.125	9.223	.007
P	38088.000	1	38088.000	3.899	.065
K	203203.125	1	203203.125	20.800	.000
N * K	.042	1	.042	.000	.998
P * K	322.667	1	322.667	.033	.858
N * P	15913.500	1	15913.500	1.629	.219
N * P * K	51987.042	1	51987.042	5.322	.034
Error	166075.500	17	9769.147		
Total	2647270.000	32			
Corrected Total	718622.000	31			

a. R Squared = .769 (Adjusted R Squared = .579)

Result: The main effects **N, P, K** have significant effect. **Block** effect is not significant, hence partial confounding is not effective.

3^2 Factorial Experiment

Case 10: To study the effects of developer strength and development time on the density of photographic film, an experiment was carried out where each of these factors with three different levels repeated 4 times. The data is obtained as follows:

DT / DS	10		14		18	
1	0 (rep 1)	2 (rep 2)	1	3	2	5
	5 (rep 3)	4 (rep 4)	4	2	4	6
2	4	6	6	8	9	10
	7	5	7	7	8	5
3	7	10	10	10	12	10
	8	7	8	7	9	8

DS: Developer strength and DT: Development time

Test for the significance of both the factors and their all possible interactions.

Here we have two treatments at 3 levels. Hence it is a 3^2 factorial experiment. Here we have two factors developers strength and developers time. The levels are: 1,2,3 for developers strength and for 10,14,18 for developers time. Each factor is replicated 4 times.

$$H_0: The\ treatments\ do\ not\ have\ significant\ effect.$$

$$H_1: The\ treatments\ have\ a\ significant\ effect.$$

The ANOVA table is given as:

Tests of Between-Subjects Effects

Dependent Variable: Density

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	238.556[a]	11	21.687	9.185	.000
Intercept	1418.778	1	1418.778	600.894	.000
DT * DS	3.278	4	.819	.347	.843
DT	22.722	2	11.361	4.812	.017
DS	198.222	2	99.111	41.976	.000
Replicate	14.333	3	4.778	2.024	.137
Error	56.667	24	2.361		
Total	1714.000	36			
Corrected Total	295.222	35			

a. R Squared = .808 (Adjusted R Squared = .720)

Result: We observe that **developers strength** and **developers time** have significant effect on photographic film. But their interaction does not have a significant effect on photographic film.