Chapter 3

Experiments with a Single Factor: The Analysis of Variance

3.2 THE ANALYSIS OF VARIANCE

The objectives of The analysis of Variance

- Study and analysis the impact of one qualitative variable or more on quantitative variable.
- Testing the equality treatment means.
- Comparability between treatment means.

Form of data

Suppose we have a treatments or different levels of a single factor that we wish to compare. The observed response from each of the a treatments is a random variable. The data would appear as in Table 3.1

Table 3.1: Form of data

Treatment (level)	**Observations**				**Total** $y_{i\cdot}$	**Mean** $\bar{y}_{i\cdot}$
1	y_{11}	y_{12}	$\cdots$	y_{1n}	$\mathbf{y_{1\cdot}}$	$\bar{\mathbf{y}}_{\mathbf{1}\cdot}$
2	y_{21}	y_{22}	$\cdots$	y_{2n}	$\mathbf{y_{2\cdot}}$	$\bar{\mathbf{y}}_{\mathbf{2}\cdot}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
a	y_{a1}	y_{a2}	$\cdots$	y_{an}	$\mathbf{y_{a\cdot}}$	$\bar{\mathbf{y}}_{\mathbf{a}\cdot}$
					$\mathbf{y_{\cdot\cdot}}$	$\bar{\mathbf{y}}_{\cdot\cdot}$

Models for the Data

We will find it useful to describe the observations from an experiment with a model. There is two forms to represent the model

- **Mean Model**

$$y_{ij} = \mu_i + \varepsilon_{ij}, \quad i = 1,2, \dots, a, \quad j = 1,2, \dots, n \qquad (3.1)$$

y_{ij}: The ijth observation of response variable

μ_i: The mean of the ith factor level or treatment

ε_{ij}: The random errors

- ## Effect Model

Equation 3.1 is called the means model. An alternative way to write a model for the data is to define the model

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

so that Equation 3.1 becomes

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij}, \ i = 1,2,\ldots,a, \quad j = 1,2,\ldots,n \qquad (3.2)$$

In this form of the model, μ is called the overall mean, and τ_i is called the ith treatment effect. Equation 3.2 is usually called the effects model.

Both the means model (3.1) and the effects model (3.2) are linear statistical models; that is, the response variable y_{ij} is a linear function of the model parameters (μ_i, μ, τ_i)

Fixed and Random Effect Model

The statistical model, Equation 3.2, describes two different situations with respect to the treatment effects.

- ***Fixed Effect:*** If we wish to test hypotheses about the treatment means, and our conclusions will apply only to the factor levels considered in the analysis. We may also wish to estimate the model parameters (μ, τ_i, σ^2)
- *Random Effect*: If the a treatments could be a random sample from a larger population of treatments.

▪ Assumptions of the Model

- The model errors ε_{ij} are assumed to be normally and independently distributed random variables with mean zero and variance σ^2 . The variance σ^2 is assumed to be constant for all levels of the factor. This implies that the observations so that, $y_{ij} \sim N(\mu + \tau_i, \sigma^2)$
- $\sum_{i=1}^{a} \tau_i = 0$

▪ Parameters Estimation of Fixed Effect Model

Table 3.2: Estimation of Model Parameters

Parameter		Estimat
Treatment Mean	μ_i	$\bar{y}_{i.} = \frac{y_{i.}}{n}$
Overall mean	μ	$\bar{y}_{..} = \frac{y_{..}}{N} = \frac{y_{..}}{an}$
Treatment effect	τ_i	$\hat{\tau}_{i.} = \bar{y}_{i.} - \bar{y}_{..}$
Variance of errors	σ^2	$\hat{\sigma}^2 = MS_E = \frac{SS_E}{N-a}$

3.3 Analysis of the Fixed Effect Model

We are interested in testing the equality of the a treatment means; that is, $\mathrm{E}(y_{ij}) = \mu + \tau_i = \mu_i.\ \mathrm{i} = 1, 2, \ldots, a$. The appropriate hypotheses are:

$$H_0: \mu_1 = \mu_2 = \cdots = \mu_a \ ,$$

$$H_1: \mu_i \neq \mu_j \text{ for at least one pair } (i, j) \qquad (3.4)$$

The treatment mean $\mu_i = \mu + \tau_i$, and $\mu = \sum_{i=1}^{a} \mu_i / a$, it implies that an other way to write the above hypotheses is in terms of the treatment effects τ_i , say: $H_0: \tau_1 = \tau_2 = \cdots = \tau_a = 0$,

$$H_1: \tau_i \neq 0 \text{ for at least one } i \qquad (3.5)$$

3.3.1 Decomposition of the Total Sum of Squares

The name analysis of variance is derived from a partitioning of total variability into its component parts. The total corrected sum of squares:

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

is used as a measure of overall variability in the data. If we divide SS_T by the appropriate number of degrees of freedom, $(an - 1 = N - 1)$ we would have the sample variance of the $y_{ij}'s$.

The total corrected sum of squares SS_T may be written.

$$\sum_{i=1}^{a}\sum_{j=1}^{n}\left(y_{ij}-\bar{y}_{..}\right)^2=\sum_{i=1}^{a}\sum_{j=1}^{n}(\bar{y}_{i.}-\bar{y}_{..})^2+\sum_{i=1}^{a}\sum_{j=1}^{n}\left(y_{ij}-\bar{y}_{i.}\right)^2 \tag{3.5}$$

Equation 3.5, can be written as another form

$$\underbrace{\sum_{i=1}^{a}\sum_{j=1}^{n}\left(y_{ij}-\bar{y}_{..}\right)^2}_{\substack{\text{total corrected sum of squares}\\ SS_T}}=\underbrace{n\sum_{j=1}^{n}(\bar{y}_{i.}-\bar{y}_{..})^2}_{\substack{\text{sum squares between the treatment}\\ SS_{Treatments}}}+\underbrace{\sum_{i=1}^{a}\sum_{j=1}^{n}\left(y_{ij}-\bar{y}_{i.}\right)^2}_{\substack{\text{sum squares within the treatment}\\ SS_E}} \tag{3.6}$$

$df \rightarrow \quad N-1 \qquad a-1 \qquad N-a$

The quantities: $\left\{MS_{Treatments}=\frac{SS_{Treatments}}{a-1},\quad MS_E=\frac{SS_E}{N-a}\right\}$ are called mean squares.

The expected values of these mean squares given by.

$$E(MS_E)=\sigma^2, \qquad E(MS_{Treatments})=\sigma^2+\frac{n\sum_{i=1}^{a}\tau_i^2}{a-1}$$

3.3.2 Statistical Analysis

The formal of test statistics used to test the hypothesis of no differences in treatment means, $H_0: \mu_1 = \mu_2 = \cdots = \mu_a$ or equivalently $H_0: \tau_1 = \tau_2 = \cdots = \tau_a = 0$, can be performed as.

$$F_0 = \frac{\frac{SS_{Treatments}}{a-1}}{\frac{SS_E}{N-a}} = \frac{MS_{Treatments}}{MS_E} \qquad (3.7)$$

is distributed as F with $(a-1, N-a)$ a degrees of freedom. Equation 3.7 is the **test statistic** for the hypothesis of no differences in treatment means.

We should reject H_0 and conclude that there are differences in the treatment means if $F_0 > F_{\alpha,\, a-1,N-a}$.

○ The mathematical formula for finding the sum of squares

Another approach is to rewrite and simplify the definitions of $SS_T, SS_{Treatments}, SS_E$ Equation 3.6, which results in

$$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{n} y_{ij}^2 - \frac{y_{..}^2}{N} \tag{3.8}$$

$$SS_{Treatments} = \frac{1}{n}\sum_{i=1}^{a} y_{i.}^2 - \frac{y_{..}^2}{N} \tag{3.9}$$

and

$$SS_E = SS_T - SS_{Treatments} \tag{3.10}$$

ANOVA table

■ TA B L E 3 . 3: The Analysis of Variance Table for the Single-Factor, Fixed Effects Model

Source of Variations	**Sum of Squares**	**Degrees of Freedom**	**Mean Squares**	F_0
Between treatments	$SS_{Treatments}$	$a-1$	$MS_{Treatments}$	$F_0 = \frac{MS_{Treatments}}{MS_E}$
Errors (Within Treatments)	SS_E	$N-a$	MS_E	
Total	SS_T	$(N-1)$		

○ EXAMPLE 3.1 :The Plasma Etching Experiment (Page 76)

An engineer wants to test four levels of RF power: 160, 180, 200, and 220 W. She decided to test five wafers at each level of RF power. The data summarized in the following table

■ **TABLE 3.1**

Etch Rate Data (in Å/min) from the Plasma Etching Experiment

Power (W)	Observations 1	2	3	4	5	Totals $y_{i.}$	Averages $\bar{y}_{i.}$
160	575	542	530	539	570	2756	551.2
180	565	593	590	579	610	2937	587.4
200	600	651	610	637	629	3127	625.4
220	725	700	715	685	710	3535	707.0
						$y_{..}$ 12,355	$\bar{y}_{..}$ 617.75

Use the analysis of variance to test $H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$ against the alternative $H_1: some\ means\ are\ different$. Apply the sums of squares in Equations 3.8, 3.9, and 3.10.

- The Answer

- **Compute Sum of Squares**

$$CF = \frac{y_{..}^2}{N} = \frac{(12355)^2}{20} = 7632301.25$$

$$Q_{To} = \sum_{i=1}^{4}\sum_{j=1}^{5} y_{ij}^2 = 575^2 + 542^2 + \cdots + 710^2 = 7704511$$

$$Q_{Tr} = \frac{1}{n}\sum_{i=1}^{4} y_{i.}^2 = \frac{1}{5}(2756^2 + 2937^2 + 3127^2 + 3535^2) = \frac{38495859}{5} = 7699171.8$$

$$SS_T = Q_{To} - \mathrm{CF} = 7704511 - 7632301.25 = 72209.75$$

$$SS_{Treatment} = Q_{Tr} - \mathrm{CF} = 7699171.8 - 7632301.25 = 66870.55$$

$$SS_E = SS_T - SS_{Treatments} = 72209.75 - 66870.55 = 5339.2$$

- **Compute F Statistics**

$$MS_{Treatments} = \frac{SS_{Treatments}}{a-1} = \frac{66870.55}{4-1} = 22209.75$$

$$MS_E = \frac{SS_E}{N-a} = \frac{5339.2}{20-4} = 333.7$$

$$F_0 = \frac{MS_{Treatments}}{MS_E} = \frac{22290.18333}{333.7} = 66.797$$

- **Steps of hypothesis test**

$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$ against $H_1: \mu_i \neq \mu_j$ for at least one pair (i, j)

$F_0 = 66.797$

$F_{\alpha, a-1, N-a} = F_{0.05,3,16} = 3.24$

Because $F_0 = 66.797 > F_{0.05,3,16} = 3.24$, we reject H_0 and conclude that the treatment means differ; that is, the RF power setting significantly affects the mean etch.

We can display ANOVA table as follows

Source of Variations	Sum of Squares	Degrees of Freedom	Mean Squares	F_0
Between treatments	66870.55	3	72209.75	66.797
Errors (Within Treatments)	5339.2	16	333.7	
Total	72209.75	19		

3.3.2 Unbalance Design

If the number of observations taken within each treatment may be different. We then say that the design is unbalanced. Then slight modifications must be made in the sum of squares formulas. Let n_i observations be taken under treatment i $(i = 1,2,\cdots,a)$ and $N = \sum_{i=1}^{a} n_i$. The manual computational formulas for SS_T and $SS_{Treatments}$ become

$$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{n_i} y_{ij}^2 - \frac{y_{..}^2}{N} \qquad (3.14)$$

and

$$SS_{Treatments} = \sum_{i=1}^{a} \frac{y_{i.}^2}{n_i} - \frac{y_{..}^2}{N} \qquad (3.15)$$

No other changes are required in the analysis of variance.

Exercise 3.16 page 134

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

Temperature	Density				
100	21.8	21.9	21.7	21.6	21.7
125	21.7	21.4	21.5	21.4	
150	21.9	21.8	21.8	21.6	21.5
175	21.9	21.7	21.8	21.4	

Does the firing temperature affect the density of the bricks? Use $\alpha = 0.05$.

The Answer

Temperature	Density					n_i	$y_{i\cdot}$	$\bar{y}_{i\cdot}$
100	21.8	21.9	21.7	21.6	21.7	5	108.7	21.74
125	21.7	21.4	21.5	21.4		4	86	21.5
150	21.9	21.8	21.8	21.6	21.5	5	108.6	21.72
175	21.9	21.7	21.8	21.4		4	86.8	21.7
						18	390.1	21.6722

$$CF = \frac{y_{\cdot\cdot}^2}{N} = \frac{(390.1)^2}{18} = 8454.3339,$$

$$Q_{To} = \sum_{i=1}^{4}\sum_{j=1}^{n_i} y_{ij}^2 = 21.8^2 + 21.9^2 + \cdots + 21.4^2 = 8454.85$$

$$Q_{Tr} = \sum_{i=1}^{4} y_{i\cdot}^2/n_i = \left(\frac{108.7^2}{5} + \frac{86^2}{4} + \frac{108.6^2}{5} + \frac{86.8^2}{4}\right) = 8454.49$$

$$\boldsymbol{SS_T} = Q_{To} - CF = 8454.85 - 8454.3339 = 0.5161,$$

$$\boldsymbol{SS_{Treatments}} = Q_{Tr} - CF = 8454.49 - 8454.3339 = 0.1561, \quad \boldsymbol{SS_E} = \boldsymbol{SS_T} - \boldsymbol{SS_{Treatments}} = 0.36$$

ANOVA table

Source of Variations	Sum of Squares	Degrees of Freedom	Mean Squares	F_0	$F_{0.05,3,14}$
Between treatments	0.1561	3	0.052033	2.0235	3.34
Errors (Within Treatments)	0.36	14	0.025714		
Total	0.5161	17			

- **Steps of hypothesis test**

$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$ against $H_1: \mu_i \neq \mu_j$ for at least one pair (i, j)

$F_0 = 2.0235$ $F_{\alpha,a-1,N-a} = F_{0.05,3,14} = 3.34$

Because $F_0 = 2.0235 < F_{0.05,3,16} = 3.34$, we accept H_0 and conclude that the all treatment means equals; that is, the temperature levels not significantly affect the density of the bricks.

HW 3.1: problems 3.12 page 131

3.3. A computer ANOVA output is shown below. Fill in the blanks. You may give bounds on the P-value.

One-way ANOVA

Source	DF	SS	MS	F	P
Factor	3	36.15	?	?	?
Error	?	?	?		
Total	19	196.04			

3.7. The tensile strength of Portland cement is being studied. Four different mixing techniques can be used economically. A completely randomized experiment was conducted and the following data were collected:

Mixing Technique	Tensile Strength (lb/in^2)			
1	3129	3000	2865	2890
2	3200	3300	2975	3150
3	2800	2900	2985	3050
4	2600	2700	2600	2765

(a) Test the hypothesis that mixing techniques affect the strength of the cement. Use $\alpha = 0.05$.