Statistical Inference

Two- Way Analysis of Variance

Contents

Two Way ANOVA

- Two Way Anova is used when there are 2 factors under study.

- Each factor can have 2 or more levels . Example: Gender and Age can be 2 factors.
  Gender with 2 levels as Male and Female
  Age with 3 levels as 18-30, 31-50 and >50

- Three hypothesis are tested.

  Factor A H0: All group means are equal
  H1: At least one mean is different from other means

  Factor B H0: All group means are equal
  H1: At least one mean is different from other means

  Interaction H0: The interaction is not significant
  H1: The interaction is significant

* For two-way ANOVA with interaction there has to be more than one observation per combination of the levels of factors.

Two Way ANOVA

Total variation is partitioned as below :

Total SS= Between Groups SS due to factor A (SSA)

+ Between Groups SS due to factor B (SSB)

+ Interaction SS due to factor A and B (SSAB)

+ Error SS (SSE)

where, SS stands for sum of squares

* SS formulae for two-way ANOVA with interaction are not specified due to their complexity.

Case Study

Background

A large company is assessing the difference in 'Satisfaction Index' of employees in Finance, Marketing and Client-Servicing departments. Experience level is also considered in the study.(<=5 years and >5 years)

Objective

To test the equality of the satisfaction index among employees of three departments (CS, Marketing, Finance) and among different experience bands.

Sample Size

Sample size: 36
Variables: satindex, dept, exp

Data Snapshot

Two Way Anova

Variables

Observations

satindex	dept	exp
75	FINANCE	lt5
56	FINANCE	lt5
62	FINANCE	gt5
66	FINANCE	gt5
58	FINANCE	gt5
58	MARKETIN	lt5
63	MARKETIN	lt5
53	MARKETIN	lt5
74	MARKETIN	lt5
77	MARKETIN	lt5
69	MARKETIN	lt5
57	MARKETIN	gt5
70	MARKETIN	gt5
68	MARKETIN	gt5
77	CS	lt5
71	CS	lt5

Columns	Description	Type	Measurement	Possible values
Satindex	Satisfaction Index	Numeric	-	Positive Values
Dept	Department	Character	MARKETING, CS, FINANCE	3
Exp	Years of Experience (grouped)	Character	lt5 = less than 5, gt5 = greater than 5	2

Two Way ANOVA

Testing equality of means in two factors.

Objective	To compare employee satisfaction index in three departments (CS, Marketing, Finance) and two experience level based groups.

Null Hypothesis

(H_{01}): Average satisfaction index is equal for 3 departments.

(H_{02}): Average satisfaction index is equal for 2 experience levels.

(H_{03}) Interaction effect(dept*exp) is not significant on satisfaction index.

The test statistic is computed for each of these null hypothesis.

Reject the null hypothesis **if p-value < 0.05**

Two Way ANOVA in R

```
# Import data
data<-read.csv("Two Way Anova.csv", header=TRUE)

# ANOVA Table
anovatable<-aov(formula=satindex~dept+exp+dept*exp,data=data)
summary(anovatable)
```

- *'aov' is the R function for ANOVA .*
- *formula specifies 'satindex' as analysis (dependent) variable and 'dept' and 'exp' as factor (independent) variables.*
- *dept*exp specifies the interaction effect.*
- *anovatable is user defined object name created to store output.*
- *summary function displays the ANOVA table output.*

Two Way ANOVA in R

```
# Output:
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
dept	2	164.2	82.11	1.679	0.204
exp	1	78.0	78.03	1.595	0.216
dept:exp	2	20.2	10.11	0.207	0.814
Residuals	30	1467.2	48.91		

Interpretation :

- *Since p-value is >0.05 for all three (dept, exp and dept*exp), do not reject H0 for all three tests. There is no significant difference in satisfaction index among 3 different departments and 2 experience levels.*
- *Also interaction effect is not significant.*

Quick Recap

Two Way Anova	• two way ANOVA is an extension of one way ANOVA when we have 2 factors in the study instead of one.
Null Hypothesis Drawing Inference	• Equality of means for levels in factor A • Equality of means for levels in factor B • No Interaction effect between 2 factors • Total sum of squares is split into 4 parts and each hypothesis is tested.

Knowledge check question

- A large retailer is testing a marketing campaign on 24 stores. 8 stores are selected randomly from each of 3 zones.
- The variable of interest is ' sales increment(%) during campaign month'. Objective is to test whether the campaign is equally effective in 3 regions. Data is given below.

NORTH	WEST	SOUTH
8	10.2	5.3
12.5	9.3	5.8
9.2	9.9	6
6.7	8.7	7.1
9.4	9.1	7
5.9	10.2	6.1
7.7	9.5	6.3
6.9	10	7.3

Is this One-way ANOVA problem or Two-way ANOVA problem?

ANSWER : One-way ANOVA

EXPLANATION : There is only one factor (zone) with 3 levels (North, West, South).