

Statistics Software Lab Report - 11

Name of the Student: Shatansh Patnaik
Roll No: 20MA20067

IIT Kharagpur
Statistics Software Lab

One Way Analysis of Variance (ANOVA)

One Way Analysis of Variance (ANOVA) is a statistical method used to compare the means of three or more groups to determine if at least one of the group means is significantly different from the others. It is commonly used when there are multiple groups or treatments and one wants to determine if there is a significant difference in means among them.

Algorithm for One Way ANOVA

The following algorithm outlines the steps for conducting a One Way ANOVA:

Algorithm 1 One Way ANOVA

- 1: **Input:** Data from k groups: X_1, X_2, \dots, X_k
 - 2: **Output:** F-statistic and p-value
 - 3: Compute the overall mean: $\bar{X} = \frac{1}{n} \sum_{i=1}^k \sum_{j=1}^{n_i} X_{ij}$
 - 4: Compute the between-group sum of squares (SSB): $SSB = \sum_{i=1}^k n_i (\bar{X}_i - \bar{X})^2$
 - 5: Compute the within-group sum of squares (SSW): $SSW = \sum_{i=1}^k \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2$
 - 6: Compute the degrees of freedom for SSB: $df_{SSB} = k - 1$
 - 7: Compute the degrees of freedom for SSW: $df_{SSW} = N - k$
 - 8: Compute the mean square for SSB: $MSB = \frac{SSB}{df_{SSB}}$
 - 9: Compute the mean square for SSW: $MSE = \frac{SSW}{df_{SSW}}$
 - 10: Compute the F-statistic: $F = \frac{MSB}{MSE}$
 - 11: Compute the p-value using the F-distribution with degrees of freedom df_{SSB} and df_{SSW}
-

ANOVA is a statistical method used to compare the means of three or more groups to determine if at least one of the group means is significantly different from the others. The algorithm takes as input the data from k groups denoted as X_1, X_2, \dots, X_k , where n_i represents the number of observations in each group X_i .

The algorithm begins by computing the overall mean \bar{X} of all the observations. It then calculates the between-group sum of squares (SSB), representing the variability between the group means, and the within-group sum of squares (SSW), representing the variability within each group. The degrees of freedom for both SSB and SSW are determined accordingly.

Next, the algorithm computes the mean square for SSB (MSB) and the mean square for SSW (MSE). These values are then used to calculate the F-statistic, which is the ratio of the variance between groups to the variance within groups.

Finally, the algorithm computes the p-value associated with the F-statistic using the F-distribution with degrees of freedom df_{SSB} and df_{SSW} . This p-value indicates the probability of observing the calculated F-statistic under the null hypothesis that there is no significant difference between the group means.

The following is the code in R for the implementation of One Way ANOVA

Code for One Way ANOVA: Problem-1 Solution

```
1 # Exercise-1:
2 # One way analysis of variance (ANOVA) is used to test the null hypothesis
  that three or more means are equal.
3 A <- c(40, 30, 50, 50, 30)
```

```

4 B <- c(60, 40, 55, 65)
5 C <- c(60, 50, 70, 65, 75, 40)
6
7 grandC <- c(A, B, C)
8 meanA <- mean(A)
9 meanB <- mean(B)
10 meanC <- mean(C)
11 meanGrandC <- mean(grandC)
12
13 findSumOfSquaresBetween <- function(A, B, C, meanGrandC) {
14   ans <- length(A) * (mean(A) - meanGrandC)^2 + length(B) * (mean(B) -
15     meanGrandC)^2 + length(C) * (mean(C) - meanGrandC)^2
16   return (ans)
17 }
18
19 findSumOfSquaresWithin <- function(A, B, C, meanA, meanB, meanC) {
20   ans <- sum((A - meanA)^2) + sum((B - meanB)^2) + sum((C - meanC)^2)
21   return (ans)
22 }
23
24 SSW <- findSumOfSquaresWithin(A, B, C, meanA, meanB, meanC)
25 SSB <- findSumOfSquaresBetween(A, B, C, meanGrandC)
26
27 dfW <- length(grandC) - 3
28 dfB <- 3 - 1
29
30 MSW <- SSW / dfW
31 MSB <- SSB / dfB
32
33 F <- MSB / MSW
34 F1 <- 2.81
35 F2 <- 3.89
36
37 if(F>F1){
38   print("Reject the Hypothesis that all means are significantly similar at
39     significance level 0.1")
40 }else{
41   print("Accept the Hypothesis that all means are significantly similar at
42     significance level 0.1")
43 }
44
45 if(F>F2){
46   print("Reject the Hypothesis that all means are significantly similar at
47     significance level 0.05")
48 }else{
49   print("Accept the Hypothesis that all means are significantly similar at
50     significance level 0.05")
51 }
52
53 # We need to calculate which group contributes the most to the difference
54 t1 <- 1.782
55 t2 <- 2.179

```

```

51
52 LSD <-t2* sqrt(MSW * (1/length(A) + 1/length(B) + 1/length(C)))
53
54 # Comparision between Brand A and Brand B
55 if(abs(meanA - meanB) > LSD){
56   print("Reject the Hypothesis that Brand A and Brand B are significantly
57         similar at significance level 0.05")
58 }else{
59   print("Accept the Hypothesis that Brand A and Brand B are significantly
60         similar at significance level 0.05")
61 }
62
63 # Comparision between Brand A and Brand C
64 if(abs(meanA - meanC) > LSD){
65   print("Reject the Hypothesis that Brand A and Brand C are significantly
66         similar at significance level 0.05")
67 }else{
68   print("Accept the Hypothesis that Brand A and Brand C are significantly
69         similar at significance level 0.05")
70 }
71
72 # Comparision between Brand B and Brand C
73 if(abs(meanB - meanC) > LSD){
74   print("Reject the Hypothesis that Brand B and Brand C are significantly
75         similar at significance level 0.05")
76 }else{
77   print("Accept the Hypothesis that Brand B and Brand C are significantly
78         similar at significance level 0.05")
79 }

```

Two Way Analysis of Variance (ANOVA)

Two Way Analysis of Variance (ANOVA) is a statistical method used to analyze the effects of two categorical independent variables (factors) on a continuous dependent variable. It allows researchers to examine not only the main effects of each factor but also their interaction effect.

Algorithm for Two Way ANOVA

The following algorithm outlines the steps for conducting Two Way ANOVA:

The implementation of the Two Way ANOVA algorithm involves several steps. First, the overall mean of the dataset is calculated. Then, the sums of squares for the two factors (Factor A and Factor B) and their interaction are computed. From these sums of squares, the degrees of freedom are determined. Next, the mean squares for each factor and their interaction are calculated. Finally, F-statistics are computed for each factor and their interaction, along with their associated p-values using the F-distribution. This implementation provides researchers with a systematic approach to assess the significance of main effects and interactions in Two Way ANOVA, helping to elucidate the relationships between categorical variables and a continuous outcome.

The following is the implementation of Two Way ANOVA in R:

Algorithm 2 Two Way ANOVA

- 1: **Input:** Data from r levels of factor A and c levels of factor B: X_{ij}
 - 2: **Output:** F-statistic and p-values for main effects and interaction
 - 3: Calculate the overall mean: $\bar{X} = \frac{1}{rc} \sum_{i=1}^r \sum_{j=1}^c X_{ij}$
 - 4: Compute the sums of squares for factor A (SSA), factor B (SSB), and their interaction (SSAB)
 - 5: Compute the degrees of freedom for SSA, SSB, and SSAB
 - 6: Calculate the mean squares for SSA (MSA), SSB (MSB), and SSAB (MSAB)
 - 7: Compute the F-statistics for factor A ($F_A = \frac{MSA}{MSE}$), factor B ($F_B = \frac{MSB}{MSE}$), and their interaction ($F_{AB} = \frac{MSAB}{MSE}$)
 - 8: Compute the p-values associated with F_A , F_B , and F_{AB} using the F-distribution
-

Code for Two Way ANOVA: Problem-2 and Problem-3

```
1 # Exercise-2:
2 yield <- matrix(c(8, 10, 6, 8, 3, 4, 5, 4, 7, 8, 6, 7), nrow = 4, byrow =
   FALSE)
3 colnames(yield) <- c("A", "B", "C")
4 rownames(yield) <- c("alpha", "beta", "gamma", "delta")
5
6 rowMean <- apply(yield, 1, mean)
7 colMean <- apply(yield, 2, mean)
8
9 grandMean <- mean(yield)
10 SSTotal <- sum((yield - grandMean)^2)
11
12 SSRows <- sum((rowMean - grandMean)^2) * 3 # 3 is the number of columns
13 SSCols <- sum((colMean - grandMean)^2) * 4 # 4 is the number of rows
14
15 SSError <- SSTotal - SSRows - SSCols
16
17 dfRow <- 3 # No. of rows - 1
18 dfCol <- 2 # No. of columns - 1
19 dfError <- (3 - 1) * (4 - 1)
20
21 MSRows <- SSRows / dfRow
22 MSCols <- SSCols / dfCol
23 MSError <- SSError / dfError
24
25 FRows <- MSRows / MSError
26 FCols <- MSCols / MSError
27
28 FCriticalCols <- 5.14
29 FCriticalRows <- 4.76
30
31 if (FRows > FCriticalRows) {
32   print("Reject the Hypothesis that the average fertilizer effects are
   significantly similar at significance level 0.05")
33 } else {
34   print("Accept the Hypothesis that the average fertilizer effects are
   significantly similar at significance level 0.05")
}
```

```

35 }
36
37 if (FCols > FCriticalCols) {
38     print("Reject the Hypothesis that the average yields are significantly
39           similar at significance level 0.05")
40 } else {
41     print("Accept the Hypothesis that the average yields are significantly
42           similar at significance level 0.05")
43 }
44
45 # Exercise-3:
46 clarity <- matrix(c(28.30, 42.38, 33.33, 40.42), nrow = 2, byrow = FALSE)
47 colnames(clarity) <- c("4", "6")
48 rownames(clarity) <- c("1.5", "2.5")
49
50 rowMean <- apply(clarity, 1, mean)
51 colMean <- apply(clarity, 2, mean)
52
53 grandMean <- mean(clarity)
54 SSTotal <- sum((clarity - grandMean)^2)
55
56 SSRows <- sum((rowMean - grandMean)^2) * 3 # 3 is the number of columns
57 SSCols <- sum((colMean - grandMean)^2) * 4 # 4 is the number of rows
58
59 SSInteraction <- SSTotal - SSRows - SSCols
60
61 dfRow <- nrow(clarity) - 1 # No. of rows - 1
62 dfCol <- ncol(clarity) - 1 # No. of columns - 1
63 dfInteraction <- dfRow * dfCol
64 dfTotal <- nrow(clarity) * ncol(clarity) - 1
65
66 MSRows <- SSRows / dfRow
67 MSCols <- SSCols / dfCol
68 MSInteraction <- SSInteraction / dfInteraction
69
70 FRows <- MSRows / MSInteraction
71 FCols <- MSCols / MSInteraction
72
73 FCritical <- 7.71
74
75 if (FRows > FCritical) {
76     print("Reject the Hypothesis that the means along rows are significantly
77           similar at significance level 0.05")
78 } else {
79     print("Accept the Hypothesis that the means along rows are significantly
80           similar at significance level 0.05")
81 }
82
83 if (FCols > FCritical) {
84     print("Reject the Hypothesis that the means along columns are
85           significantly similar at significance level 0.05")
86 } else {

```

```

82     print("Accept the Hypothesis that the means along columns are
      significantly similar at significance level 0.05")
83 }
84
85 if(MSInteraction > 0 && FRows < FCritical && FCols < FCritical){
86     print("Reject the Hypothesis that there is an interaction between rows and
      columns at significance level 0.05")
87 }else{
88     print("Accept the Hypothesis that there is an interaction between rows and
      columns at significance level 0.05")
89 }

```

Latin Square Design

Latin Square Design is a experimental design method used in statistics to control variation and reduce bias in experiments involving three factors. It is particularly useful in situations where there are two nuisance factors which need to be controlled. Latin Square Design ensures that each level of every factor appears exactly once in each row and column, thus minimizing the effects of extraneous variables.

Algorithm for Latin Square Design

The following algorithm outlines the steps for creating a Latin Square Design:

Algorithm 3 Latin Square Design

- 1: **Input:** Number of treatments n
 - 2: **Output:** Latin Square of size $n \times n$
 - 3: Initialize an $n \times n$ matrix with zeros
 - 4: **for** $i = 1$ to n **do**
 - 5: **for** $j = 1$ to n **do**
 - 6: Set $k = (i + j - 1) \bmod n + 1$
 - 7: Assign treatment k to position (i, j) in the matrix
 - 8: **end for**
 - 9: **end for**
-

Latin Square Design is a special experimental design used in statistics to control variation and reduce bias in experiments involving three factors. It arranges treatments in a square grid so that each treatment occurs once in each row and once in each column. Analysis of Variance (ANOVA) for Latin Square Design is used to determine if there are significant differences between the treatments while controlling for the row and column effects.

Algorithm for ANOVA for Latin Square Design

The following algorithm outlines the steps for conducting ANOVA for Latin Square Design:

This algorithm enables researchers to assess the significance of treatments while accounting for the row and column effects in Latin Square Design experiments.

Here is the implementation of Analysis of Variance of Three Way and Four Way ANOVA in R (also known as Latin Square Design and Graeco Latin Square Design Methods)

Algorithm 4 ANOVA for Latin Square Design

- 1: **Input:** Data from Latin Square Design: X_{ijl}
 - 2: **Output:** F-statistic and p-value
 - 3: Calculate the overall mean: $\bar{X} = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \sum_{l=1}^n X_{ijl}$
 - 4: Compute the sum of squares for treatments (SST): $SST = \sum_{l=1}^n (\sum_{i=1}^n \sum_{j=1}^n X_{ijl})^2 / n - \bar{X}^2$
 - 5: Compute the sum of squares for rows (SSR): $SSR = \sum_{i=1}^n (\sum_{l=1}^n \sum_{j=1}^n X_{ijl})^2 / n - \bar{X}^2$
 - 6: Compute the sum of squares for columns (SSC): $SSC = \sum_{j=1}^n (\sum_{l=1}^n \sum_{i=1}^n X_{ijl})^2 / n - \bar{X}^2$
 - 7: Compute the error sum of squares (SSE): $SSE = SST - SSR - SSC$
 - 8: Compute the degrees of freedom for SST, SSR, SSC, and SSE
 - 9: Calculate the mean squares for treatments (MST), rows (MSR), columns (MSC), and error (MSE)
 - 10: Compute the F-statistic for treatments, rows, and columns
 - 11: Compute the p-values associated with the F-statistics using the F-distribution
-

Code for Problem-4 and Problem-5

```
1  # Exercise-4:
2  data <- matrix(c(10, 14, 7, 8,
3                  7, 18, 11, 8,
4                  5, 10, 11, 9,
5                  10, 10, 12, 14), nrow = 4, byrow = TRUE)
6
7  globalMean <- mean(data)
8  data <- data - globalMean
9  n <- 4
10
11  treatments <- c(-11, -4, 12, 3)
12  rowSums <- apply(data, 1, sum)
13  colSums <- apply(data, 2, sum)
14
15  SSRows <- sum(rowSums^2) / 4
16  SSCols <- sum(colSums^2) / 4
17  SSTreatments <- sum(treatments^2) / 4
18  SSTotal <- sum(data^2)
19  SSError <- SSTotal - SSRows - SSCols - SSTreatments
20
21  dfRows <- n - 1
22  dfCols <- n - 1
23  dfTreatments <- n - 1
24  dfTotal <- n^2 - 1
25  dfError <- dfTotal - dfRows - dfCols - dfTreatments
26
27  MSRows <- SSRows / dfRows
28  MSCols <- SSCols / dfCols
29  MSTreatments <- SSTreatments / dfTreatments
30  MSError <- SSError / dfError
31
32  FRows <- MSRows / MSError
33  FCols <- MSCols / MSError
```



```

34 | FTreatments <- MSTreatments / MSError
35 |
36 | FCritical <- 4.76
37 |
38 | #Print everything obtained till now
39 | cat("The data is: ", data, "\n")
40 | cat("The sum of squares along rows is: ", SSRows, "\n")
41 | cat("The sum of squares along columns is: ", SSCols, "\n")
42 | cat("The sum of squares of treatments is: ", SSTreatments, "\n")
43 | cat("The sum of squares of error is: ", SSErrors, "\n")
44 | cat("The degrees of freedom along rows is: ", dfRows, "\n")
45 | cat("The degrees of freedom along columns is: ", dfCols, "\n")
46 | cat("The degrees of freedom of treatments is: ", dfTreatments, "\n")
47 | cat("The degrees of freedom of error is: ", dfError, "\n")
48 | cat("The mean square along rows is: ", MSRows, "\n")
49 | cat("The mean square along columns is: ", MSCols, "\n")
50 | cat("The mean square of treatments is: ", MSTreatments, "\n")
51 | cat("The mean square of error is: ", MSErrors, "\n")
52 | cat("The F value along rows is: ", FRows, "\n")
53 | cat("The F value along columns is: ", FCols, "\n")
54 | cat("The F value of treatments is: ", FTreatments, "\n")
55 |
56 | if(FRows > FCritical){
57 |   print("Reject the Hypothesis that the means along rows are significantly
58 |     similar at significance level 0.05")
59 | }else{
60 |   print("Accept the Hypothesis that the means along rows are significantly
61 |     similar at significance level 0.05")
62 | }
63 |
64 | if(FCols > FCritical){
65 |   print("Reject the Hypothesis that the means along columns are
66 |     significantly similar at significance level 0.05")
67 | }else{
68 |   print("Accept the Hypothesis that the means along columns are
69 |     significantly similar at significance level 0.05")
70 | }
71 |
72 | if(FTreatments > FCritical){
73 |   print("Reject the Hypothesis that the means of treatments are
74 |     significantly similar at significance level 0.05")
75 | }else{
76 |   print("Accept the Hypothesis that the means of treatments are
77 |     significantly similar at significance level 0.05")
78 | }
79 |
80 | # Exercise - 5:
81 | data <- matrix(c(11, 10, 14, 8,
82 |                 8, 12, 10, 12,
83 |                 9, 11, 7, 15,
84 |                 9, 8, 18, 6), nrow = 4, byrow = TRUE)

```

```

80
81 globalMean <- mean(data)
82 data <- data - globalMean
83 n <- 4
84
85 treatments <- c(-7, -11, 14, 4)
86 workplace <- c(3, 4, 2, -1)
87 rowSums <- apply(data, 1, sum)
88 colSums <- apply(data, 2, sum)
89
90 SSRows <- sum(rowSums^2) / 4
91 SSCols <- sum(colSums^2) / 4
92 SSTreatments <- sum(treatments^2) / 4
93 SSWorkplace <- sum(workplace^2) / 4
94 SSTotal <- sum(data^2)
95 SSError <- SSTotal - SSRows - SSCols - SSTreatments - SSWorkplace
96
97 dfRows <- n - 1
98 dfCols <- n - 1
99 dfTreatments <- n - 1
100 dfWorkplace <- n - 1
101 dfTotal <- n^2 - 1
102 dfError <- dfTotal - dfRows - dfCols - dfTreatments - dfWorkplace
103
104 MSRows <- SSRows / dfRows
105 MSCols <- SSCols / dfCols
106 MSTreatments <- SSTreatments / dfTreatments
107 MSWorkplace <- SSWorkplace / dfWorkplace
108 MSError <- SSError / dfError
109
110 FRows <- MSRows / MSError
111 FCols <- MSCols / MSError
112 FTreatments <- MSTreatments / MSError
113 FWorkplace <- MSWorkplace / MSError
114
115 FCritical <- 9.28
116
117 #Print everything obtained till now
118 cat("The data is: ", data, "\n")
119 cat("The sum of squares along rows is: ", SSRows, "\n")
120 cat("The sum of squares along columns is: ", SSCols, "\n")
121 cat("The sum of squares of treatments is: ", SSTreatments, "\n")
122 cat("The sum of squares of workplace is: ", SSWorkplace, "\n")
123 cat("The sum of squares of error is: ", SSError, "\n")
124 cat("The degrees of freedom along rows is: ", dfRows, "\n")
125 cat("The degrees of freedom along columns is: ", dfCols, "\n")
126 cat("The degrees of freedom of treatments is: ", dfTreatments, "\n")
127 cat("The degrees of freedom of workplace is: ", dfWorkplace, "\n")
128 cat("The degrees of freedom of error is: ", dfError, "\n")
129 cat("The mean square along rows is: ", MSRows, "\n")
130 cat("The mean square along columns is: ", MSCols, "\n")
131 cat("The mean square of treatments is: ", MSTreatments, "\n")

```

```

132 cat("The mean square of workplace is: ", MSWorkplace, "\n")
133 cat("The mean square of error is: ", MSEError, "\n")
134 cat("The F value along rows is: ", FRows, "\n")
135 cat("The F value along columns is: ", FCols, "\n")
136 cat("The F value of treatments is: ", FTreatments, "\n")
137 cat("The F value of workplace is: ", FWorkplace, "\n")
138
139 if(FRows > FCritical){
140     print("Reject the Hypothesis that the means along rows are significantly
          similar at significance level 0.05")
141 }else{
142     print("Accept the Hypothesis that the means along rows are significantly
          similar at significance level 0.05")
143 }
144
145 if(FCols > FCritical){
146     print("Reject the Hypothesis that the means along columns are
          significantly similar at significance level 0.05")
147 }else{
148     print("Accept the Hypothesis that the means along columns are
          significantly similar at significance level 0.05")
149 }
150
151 if(FTreatments > FCritical){
152     print("Reject the Hypothesis that the means of treatments are
          significantly similar at significance level 0.05")
153 }else{
154     print("Accept the Hypothesis that the means of treatments are
          significantly similar at significance level 0.05")
155 }
156
157 if(FWorkplace > FCritical){
158     print("Reject the Hypothesis that the means of workplace are significantly
          similar at significance level 0.05")
159 }else{
160     print("Accept the Hypothesis that the means of workplace are significantly
          similar at significance level 0.05")
161 }

```