**Introduction:**

ANOVA is a statistical technique that is used for analyzing the differences in means in three or more independent groups. It compares the variance between-groups means to the variance within-groups. However, ANOVA is conducted using the same five step approach.

ANOVA Hypotheses

* Null hypothesis: Groups means are equal (no variation in means of groups)  
  H0: μ1=μ2=…=μp
* Alternative hypothesis: At least, one group mean is different from other groups  
  H1: All μ are not equal

Null hypothesis is tested using the F-test for all groups.

**Objectives:**

1. Perform analysis of variance by hand
2. Interpret results of analysis of variance tests
3. Identify the hypothesis testing procedure based on type of outcome variable and number of samples

**The ANOVA technique:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Group 1** | **Group 2** | **Group 3** | **Group 4** |
| **Sample Size** | n1 | n2 | n3 | n4 |
| **Sample Mean** |  |  |  |  |
| **Sample Standard Deviation** | s1 | s2 | s3 | s4 |

**Test Statistics For ANOVA**

The decision rule for the F test in ANOVA is depends on the level of significance and the degrees of freedom. The F statistic has two degrees of freedom. Moreover, these are denoted df1 and df2, and called the numerator and denominator degrees of freedom, respectively. The degrees of freedom are defined as follows:

df1 = k-1 and df2=N-k

The critical value is found in a table of probability values for the F distribution with (degrees of freedom) df1 = k-1, df2=N-k. Hence, the decision rule is: Reject H0 if Fcal > Ftab

**Calculations:**

**Sample Dataset:** There are four variables and each variable has 5 values.

|  |  |  |  |
| --- | --- | --- | --- |
| **p1** | **p2** | **p3** | **p4** |
| 89 | 93 | 89 | 81 |
| 89 | 92 | 88 | 78 |
| 88 | 94 | 89 | 81 |
| 78 | 89 | 93 | 92 |
| 78 | 88 | 90 | 82 |

We will run the ANOVA using the five-step approach.

**Step 1.** Compute the Summary Statistics.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *Groups* | *Count* | *Sum* | *Average* | *Variance* |
| p1 | 5 | 422 | 84.4 | 34.3 |
| p2 | 5 | 456 | 91.2 | 6.7 |
| p3 | 5 | 449 | 89.8 | 3.7 |
| p4 | 5 | 414 | 82.8 | 28.7 |

**Step 2.** Calculate the group means and the overall mean.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| group means | 84.4 | 91.2 | 89.8 | 82.8 |
| overall means | 87.05 |  |  |  |

**Step 3.** Calculate SSR : nΣ(Xj – X..)2

|  |  |  |
| --- | --- | --- |
| Between Groups | |  |
| SSR | 249.35 |  |
| df (number of groups - 1) | 3 |  |
| MS | 83.11666667 |  |

|  |  |  |
| --- | --- | --- |
| Within Groups |  |  |
| SSR | 293.6 |  |
| df (number of observations - number of groups) | 16 |  |
| MS | 18.35 |  |

Calculate SSE: Σ(Xij – Xj)2

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 172.3125 | 112.9125 | 52.6125 | 205.1125 |
| SSE | 542.95 |  |  |  |

In order to find out the critical value of F we need degrees of freedom, df1=k-1 and df2=N-k.

**Step 4.** Compute the test statistic.

To organize our computations, to fill the ANOVA table. In order to compute the sums of squares we must first compute the sample means for each group and the overall mean.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Fill the ANOVA table |  |  |  |  |  |  |
| *Source of Variation* | *SS* | *df* | *MS* | *F* | *P-value* | *F crit* |
| Between Groups | 249.35 | 3 | 83.11667 | 4.529519 | 0.017567 | 3.238872 |
| Within Groups | 293.6 | 16 | 18.35 |  |  |  |
|  |  |  |  |  |  |  |
| Total | 542.95 | 19 |  |  |  |  |

**Step 5.** Conclusion

We reject H0 because 4.52 > 3.23. We have statistically significant evidence at α=0.05 to show that there is a difference in mean between the four groups. Hence, observed value of F is greater than the value in the F table, then we can reject the null hypothesis with 95 percent confidence.