Statistical Inference

Post Hoc analysis in ANOVA

Contents

Post Hoc testing in ANOVA

- Post-hoc (Latin, meaning "after this") means analyzing results further after main analysis. They are often based on a **familywise error rate -** it is the probability of making at least one Type I Error.

- Post-hoc pairwise comparisons are commonly performed after significant effects have been found when there are three or more levels in a factor.

- After an ANOVA, you may know that the means of your response variable differ significantly across your factors, but you do not know which pairs of the factor levels are significantly different from each other. At this point, you can conduct pairwise comparisons.

- **Post hoc tests:**
  - Pairwise t tests (Not recommended)
  - Pairwise t tests using the Bonferroni adjustment
  - Tukey test

Case Study

To execute a Post Hoc analysis in ANOVA in R, we shall consider this case as an example.

Background

A company has recorded aptitude scores for three groups of employees.

Objective

To test whether there is no significant difference in the Mean Aptitude score of three groups of employees.

Sample Size

Sample size: 24
Variables: aptscore, group

Data Snapshot

Variables

Observations

aptscore	group
34	GrI
28	GrI
41	GrI
31	GrI
39	GrI
44	GrI
25	GrI
20	GrI
27	GrII
31	GrII
33	GrII

Columns	Description	Type	Measurement	Possible values
aptscore	Aptitude score	Numeric	-	Positive value
group	Group of employees	character	GrI,GrII,GrIII	3

ONE WAY ANOVA

Testing **equality** of more than two means

Objective	To test the **null hypothesis** that **means** scores **are same**

Null Hypothesis (H_0): **Mean Aptitude score of three groups are equal**

$$\mu_1 = \mu_2 = \mu_3$$

Alternate Hypothesis (H_1): At least one group mean is different than other

Test Statistic	**Test statistic is based on F distribution.**
Decision Criteria	Reject the null hypothesis **if p-value < 0.05**

ANOVA in R

```
# Import data
data<-read.csv("Post Hoc Tests-Anova.csv", header=TRUE)

# Anova table
anova<-aov(formula=aptscore~group,data=data)
summary(anova)
```

- *'aov' is the R function for ANOVA .*
- *anova is user defined object name created to store output.*
- *summary function displays the ANOVA table output.*

ANOVA in R

```
# Output:
```

```
            Df Sum Sq Mean Sq F value Pr(>F)
group        2  501.7  250.87   4.832 0.0188 *
Residuals   21 1090.2   51.92
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Interpretation :

- *P-value<0.05,reject h0. There is significant difference in mean aptitude scores of three groups of employees.*
- *Now ,we will do the pairwise testing to identify which pairs are having difference in aptitude score.*

Pairwise t tests

```
pairwise.t.test(data$aptscore,data$group,p.adj="none")
```

- *pairwise.t.test is the R function for pairwise comparison .*
- *p.adj =none specifies no adjustments are used.*

```
# Output:
```

```
        Pairwise comparisons using t tests with pooled SD

data:  data$aptscore and data$group

      GrI    GrII
GrII  0.6082 -
GrIII 0.0261 0.0083
```

Interpretation :

- *P-value<0.05 for group I and III and also for group II and III. Aptitude test score is significantly different between Group I and III and Group II and III.*

Why Bonferroni Adjustment is needed ?

- The Bonferroni correction is used to limit the **possibility of getting a statistically significant result** when testing multiple hypotheses. It's needed because the more tests you run, the more likely you are to get a significant result. The correction lowers the area where you can reject the null hypothesis. In other words, it makes your p-value smaller.

- **Example :** Imagine looking for the Ace of Clubs in a deck of cards: if you pull one card from the deck, the odds are pretty low (1/52) that you'll get the Ace of Clubs. Try again (and try perhaps 50 times), you'll probably end up getting the Ace. The same principal works with hypothesis testing: the more simultaneous tests you run, the more likely you'll get a "significant" result. Let's say you were running 50 tests simultaneously with an alpha level of 0.05. The probability of observing at least one significant event due to chance alone is:
- P (significant event) = 1 – P(no significant event)
- = 1 – (1-0.05)^50 = 0.92.
- That's almost certain (92%) that you'll get at least one significant result.

Pairwise t tests using Bonferroni adjustment

```
pairwise.t.test(data$aptscore,data$group,p.adj="bonf")
```

- *pairwise.t.test is the R function for pairwise comparison .*
- *p.adj =bonf specifies Bonferroni adjustments are used.*

```
# Output:
        Pairwise comparisons using t tests with pooled SD

data:  data$aptscore and data$group

      GrI   GrII
GrII  1.000 -
GrIII 0.078 0.025

P value adjustment method: bonferroni
```

Interpretation :

- *P-value<0.05 for group II and III. Aptitude test score is significantly different for Group II and III.*

Get an Edge

```
#pairwise t tests with no adjustment
pairwise.t.test(data$aptscore, data$group, p.adj="none")
```

```
        Pairwise comparisons using t tests with pooled SD

data:  data$aptscore and data$group

      GrI    GrII
GrII  0.6082 -
GrIII 0.0261 0.0083
```

```
#Bonferroni adjustment multiples p value by number of
#comparisons k(here k=3).Note that maximum p value can be one.
pairwise.t.test(data$aptscore, data$group, p.adj="bonf")
```

```
        Pairwise comparisons using t tests with pooled SD

data:  data$aptscore and data$group

      GrI   GrII
GrII  1.000 -
GrIII 0.078 0.025

P value adjustment method: bonferroni
```

Tukey test

The purpose of the Tukey's test is to figure out which groups in your sample differ. It uses the "Honest Significant Difference", a number that represents the distance between groups, to compare every mean with every other mean.

```
TukeyHSD(anova,"group")
```

- *TukeyHSD test is used for pairwise comparison .*
- *anova is the object created in aov funvtion.*
- *group is the factor variable.*

Tukey test

Output:

```
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = aptscore ~ group, data = data)

$`group`
              diff       lwr        upr     p adj
GrII-GrI     1.875  -7.20576 10.9557603 0.8622638
GrIII-GrI   -8.625 -17.70576  0.4557603 0.0646166
GrIII-GrII -10.500 -19.58076 -1.4192397 0.0216622
```

Interpretation :

- *P-value<0.05 for group II and III. Aptitude test score is significantly different for Group II and III.*

Quick Recap

Post Hoc Analysis	• Post-hoc pairwise comparisons are commonly performed after significant effects have been found when there are three or more levels of a factor.
Post Hoc tests	• Pairwise t tests • Pairwise t test using Bonferroni adjustments • Tukey test