**The 3 Testers – Tukey, Scheffe & Bonferroni**

In this blog post I will provide fully worked examples of:

* Tukey HSD test
* Scheffe method
* Bonferroni method

Consider the following:

*## set groups with values*

group1 <- c(5,5,4,4,3)

group2 <- c(5,4,4,3,3)

group3 <- c(4,3,2,2,1)

*## set-up groups dataframe*

groups <- data.frame( group1, group2, group3 )

*## visualise the data*

boxplot(groups)

***insert boxplot here***

*## stack function (take matrix and stack it back into a vector of name-value pairs)*

*## i.e. go from table view to column view*

sgroups <- stack(groups)

*## names are values & indicators (indicators are the column headings from the original matirx table)*

names(sgroups)

head(sgroups)

*## 1-way test analysis of the means*

oneway.test( values ~ ind, var.equal=TRUE, data=sgroups)

One-way analysis of means

data: values and ind

F = 4.963, num df = 2, denom df = 12, p-value = 0.02687

*## analysis of variance*

model2 <- aov( values ~ ind, data=sgroups)

summary(model2)

Df Sum Sq Mean Sq F value Pr(>F)

ind 2 8.933 4.467 4.963 0.0269 \*

Residuals 12 10.800 0.900

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

The Tukey HSD (honest significant difference) test is a single-step, [multi-comparison](https://en.wikipedia.org/wiki/Multiple_comparison) [statistical test](https://en.wikipedia.org/wiki/Statistical_test). This test is typically used to ascertain whether means are significantly different from each other, via comparing all possible pairs of [means](https://en.wikipedia.org/wiki/Sample_mean).

The Tukey HSD test methodically compares the means of every treatment to the means of all other treatments. Stating it mathematically, the test applies simultaneously to the set of all pairwise comparisons

µi - µj

Ʉ i = i(1)n, j = 1(1)n

where i ≠ j

Thus, the test identifies any difference between two means whereby such difference is greater than the expected [standard error](https://en.wikipedia.org/wiki/Standard_error). Note: The Tukey HSD test is generally accepted to be quite conservative when applied to unequal sample sizes.

The formula for Tukey's test is:

 q_s = \frac{Y_A - Y_B}{SE}, 

where *Y*A is the larger of the two means being compared,

*Y*B is the smaller of the two means being compared,

SE is the [standard error](https://en.wikipedia.org/wiki/Standard_error_%28statistics%29) of the data in question.

Let us now apply the Tukey HSD test to the data under consideration:

*## Tukey honest significant difference*

TukeyHSD(model2, conf.level = 0.95)

install.packages("multcomp")

library(multcomp)

tukeyModel <- glht(model2, linfct = mcp( ind = "Tukey" ))

summary(tukeyModel)

Simultaneous Tests for General Linear Hypotheses

Multiple Comparisons of Means: Tukey Contrasts

Fit: aov(formula = values ~ ind, data = sgroups)

Linear Hypotheses:

Estimate Std. Error t value Pr(>|t|)

group2 - group1 == 0 -0.4 0.6 -0.667 0.7867

group3 - group1 == 0 -1.8 0.6 -3.000 0.0278 \*

group3 - group2 == 0 -1.4 0.6 -2.333 0.0891 .

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Adjusted p values reported -- single-step method)

An alternative to the Tukey HSD approach is the Scheffe method. The Scheffe method is also a single-step, multi-comparison procedure. Unlike the Tukey HSD, the Scheffe method is not restricted to only the pairwise comparison difference, rather it applies to the set of estimates of all possible [contrasts](https://en.wikipedia.org/wiki/Contrast_%28statistics%29) among the factor level means.

Hence, the Scheffe method is used for adjusting [significance levels](https://en.wikipedia.org/wiki/Statistical_significance) in a [linear regression](https://en.wikipedia.org/wiki/Linear_regression) analysis to account for [multiple comparisons](https://en.wikipedia.org/wiki/Multiple_comparisons). This test is commonly used in [analysis of variance](https://en.wikipedia.org/wiki/Analysis_of_variance) (ANOVA).

Herewith the application of the Scheffe method:

*## Scheffe method*

*## create array of t-values*

Tarray <- summary(tukeyModel)$test$tstat

*## display contents of T array*

as.data.frame(Tarray)

Tarray

group2 - group1 -0.6666667

group3 - group1 -3.0000000

group3 - group2 -2.3333333

*## Capture the number of columns (k) & the degrees of freedom of error (ν) from the*

*## Anova output - stored in the R memory variable called "model2" above.*

k <- model2$rank

v <- model2$df.residual

*## calculate the p-value for each pair according to the Scheffe multiple comparison method.*

*## the F distribution is evaluated using the R built-in function, pf().*

pValuesScheffe <- 1 - pf( (Tarray\*\*2) / (k-1), k-1, v)

as.data.frame(pValuesScheffe)

pValuesScheffe

group2 - group1 0.80395992

group3 - group1 0.03481543

group3 - group2 0.10596065

Another alternative, the Bonferroni method allows numerous comparison statements to be made whilst still ensuring and overall confidence coefficient is maintained.

The Bonferroni method is applied to the ANOVA case where a specific set of pairwise comparisons is selected prior to the analysis being conducted. Such set is limited but typically exceeds the set of pairwise comparisons as per the Tukey HSD test.

And now the Bonferroni method in action:

*## Bonferroni method*

*## row bind vector "differences" for comparison*

contrasts <- rbind(

"group2 - group1" = c(-1, 1, 0),

"group3 - group1" = c(-1, 0, 1),

"group3 - group2" = c(0, -1, 1)

)

*## check row binding*

contrasts

*## calculate Bonferroni simultaneous multiple comparison for all pairs.*

bonfModel <- glht(model2, linfct = mcp(ind = contrasts))

*## argument is "bonferroni" with a lowercase "b"*

summary(bonfModel, test = adjusted("bonferroni"))

*## repeat Bonferroni comparisons for fewer relevant pairs,*

*## e.g. using group1 as the control and*

*## analysing only pairs relative to group1.*

*## redefine contrast matrix as follows*

contrasts <- rbind(

"group2 - group1" = c(-1,1,0),

"group3 - group1" = c(-1,0,1)

)

*## check row binding*

contrasts

[,1] [,2] [,3]

group2 - group1 -1 1 0

group3 - group1 -1 0 1

group3 - group2 0 -1 1

*## calculate Bonferroni simultaneous multiple comparison for*

*## for the relevant q=2 pairs of contrasts relative to group1 only.*

bonfModel <- glht(model2, linfct = mcp(ind = contrasts))

*## argument is "bonferroni" with a lowercase "b"*

summary(bonfModel, test = adjusted("bonferroni"))

Simultaneous Tests for General Linear Hypotheses

Multiple Comparisons of Means: User-defined Contrasts

Fit: aov(formula = values ~ ind, data = sgroups)

Linear Hypotheses:

Estimate Std. Error t value Pr(>|t|)

group2 - group1 == 0 -0.4 0.6 -0.667 1.0000

group3 - group1 == 0 -1.8 0.6 -3.000 0.0332 \*

group3 - group2 == 0 -1.4 0.6 -2.333 0.1135

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Adjusted p values reported -- bonferroni method)

*## calcualte Bonferroni simultaneous multiple comparison for*

*## for the relevant q=2 pairs of contrasts relative to group1 only.*

bonfModel <- glht(model2, linfct = mcp(ind = contrasts))

*## argument is "bonferroni" with a lowercase "b"*

summary(bonfModel, test = adjusted("bonferroni"))

Simultaneous Tests for General Linear Hypotheses

Multiple Comparisons of Means: User-defined Contrasts

Fit: aov(formula = values ~ ind, data = sgroups)

Linear Hypotheses:

Estimate Std. Error t value Pr(>|t|)

group2 - group1 == 0 -0.4 0.6 -0.667 1.0000

group3 - group1 == 0 -1.8 0.6 -3.000 0.0221 \*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Adjusted p values reported -- bonferroni method)