ANOVA

One-way ANOVA test

- ANOVA – Analysis of Variance
- One-way ANOVA, also known as one-factor ANOVA is a test for comparing means of more than two groups
- ANOVA test hypotheses:
  - Null hypothesis: the means of the different groups are the same
  - Alternative hypothesis: At least one sample mean is not equal to the others.

Assumptions of ANOVA test

- The observations are obtained independently and randomly from the population defined by the factor levels
- The data of each factor level are normally distributed.
- These normal populations have a common variance. (**Levene's test** can be used to check this.)

How it works?

- Assume that we have 3 groups (A, B, C) to compare:
  - Compute the common variance, which is called **variance within samples** (S^2within) or **residual variance**.
  - Compute the variance between sample means as follow:
    - Compute the mean of each group
    - Compute the **variance between sample means** (S^2between)
  - Produce F-statistic as the ratio of S^2between / S^2within.

Compute one-way ANOVA

- We want to know if there is any significant difference between the average weights of plants in the 3 experimental conditions.
- Functions used
  - aov()
  - summary.aov()

R code

```
#compute analysis of variance
res <-aov(weight~group,data=mydat)

#summary of analysis
summary.aov(res)
```

```
Df Sum Sq Mean Sq F value Pr(>F)
group 2 3.766 1.8832 4.846 0.0159 *
Residuals 27 10.492 0.3886
---
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’
0.1 ‘ ’ 1
```

As the p-value is less than the significance level 0.05, we can conclude that there are significant differences between the groups highlighted with “*" in the model summary.

Multiple pairwise-comparison between the means of groups

- In one-way ANOVA test, a significant p-value indicates that some of the group means are different, but we don’t know which pairs of groups are different.

- It’s possible to perform multiple pairwise-comparison, to determine if the mean difference between specific pairs of group are statistically significant.

Tukey multiple pairwise-comparisons

- As the ANOVA test is significant, we can compute **Tukey HSD** (Tukey Honest Significant Differences)
- R function: **TukeyHSD**() for performing multiple pairwise-comparison between the means of groups.

R code

```
#Tukey HSD -multiple pairwise
comparison
TukeyHSD(res)
```

```
Tukey multiple comparisons of means 95% family-wise
confidence level
Fit: aov(formula = weight ~ group, data = mydat)
$group diff lwr upr p adj
trt1-ctrl -0.371 -1.0622161 0.3202161 0.3908711
trt2-ctrl 0.494 -0.1972161 1.1852161 0.1979960
trt2-trt1 0.865 0.1737839 1.5562161 0.0120064
```

It can be seen from the output, that only the difference between trt2 and trt1 is significant with an adjusted p-value of 0.012.

Checking ANOVA Assumptions

- Check the homogeneity of variance assumption
  - The **residuals versus fits plot** can be used to check the homogeneity of variances.

```
#checking homogeneity of variance
plot(res,1)
```

  - Levene's test

```
#Levene's Test
library(car)
leveneTest(weight~group,mydat)
```

Checking ANOVA Assumptions (contd.)

- Check the normality assumption
  - **Normality plot of residuals**. In this plot, the quantiles of the residuals are plotted against the quantiles of the normal distribution.

```
#checking homogeneity of variance
plot(res,2)
```

  - **Shapiro-Wilk test** on ANOVA residuals

```
#Extract the residuals
res_resi <- residuals(res)
shapiro.test(res_resi)
```

Non-parametric alternative to one-way ANOVA test

- A non-parametric alternative to one-way ANOVA is **Kruskal-Wallis rank sum test**, which can be used when ANOVA assumptions are not met.

```
kruskal.test(weight~group,mydat)
```

Reference

- http://www.sthda.com/english/wiki/one-way-anova-test-in-r