

# Basic Statistics

yingying

January 10, 2019

## One-Way ANOVA Test in R

The one-way analysis of variance (ANOVA), also known as one-factor ANOVA, is an extension of independent two-samples t-test for comparing means in a situation where there are more than two groups. In one-way ANOVA, the data is organized into several groups base on one single grouping variable (also called factor variable). This tutorial describes the basic principle of the one-way ANOVA test and provides practical anova test examples in R software.

ANOVA test hypotheses:

Null hypothesis: the means of the different groups are the

Alternative hypothesis: At least one sample mean is not equ

Assumptions of ANOVA test Here we describe the requirement for ANOVA test. ANOVA test can be applied only when:

The observations are obtained independently and randomly fr

The data of each factor level are normally distributed.

These normal populations have a common variance. (Levene's

## Compute one-way ANOVA test

```
##              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
```

## Multiple pairwise-comparison between the means of groups

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. The function `TukeyHSD()` takes the fitted ANOVA as an argument.

```
## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = weight ~ group, data = my_data)
##
## $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	

## Multiple comparisons using multcomp package

It's possible to use the function `glht()` [in multcomp package] to perform multiple comparison procedures for an ANOVA. `glht` stands for general linear hypothesis tests. The simplified format is as follow:  
`glht(model, lincft)`

`model`: a fitted model, for example an object returned by `ad`  
`lincft()`: a specification of the linear hypotheses to be tested

Use `glht()` to perform multiple pairwise-comparisons for a one-way ANOVA:

```
## Loading required package: mvtnorm  
## Loading required package: survival  
## Loading required package: TH.data  
## Loading required package: MASS  
##  
## Attaching package: 'MASS'
```

## Pairwise t-test

The function `pairwise.t.test()` can be also used to calculate pairwise comparisons between group levels with corrections for multiple testing.

```
##  
## Pairwise comparisons using t tests with pooled SD  
##  
## data: my_data$weight and my_data$group  
##  
##      ctrl  trt1  
## trt1 0.194 -  
## trt2 0.132 0.013  
##  
## P value adjustment method: BH
```

## Check ANOVA assumptions: test validity?

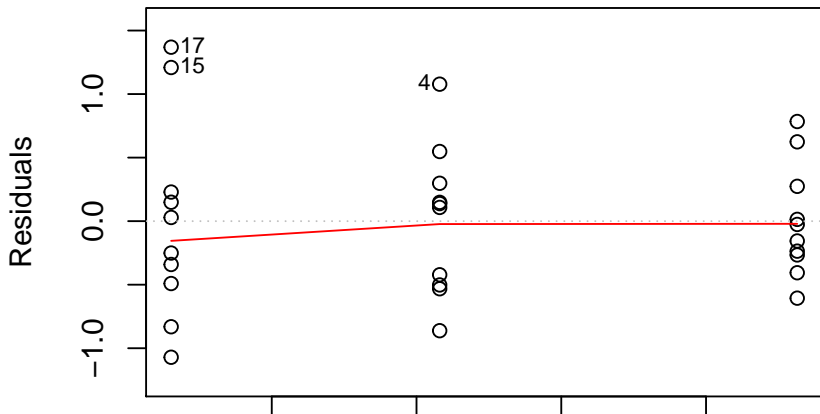
The ANOVA test assumes that, the data are normally distributed and the variance across groups are homogeneous. We can check that with some diagnostic plots.

## Check the homogeneity of variance assumption

The residuals versus fits plot can be used to check the homogeneity of variances.

In the plot below, there is no evident relationships between residuals and fitted values (the mean of each groups), which is good. So, we can assume the homogeneity of variances.

Residuals vs Fitted





## Relaxing the homogeneity of variance assumption

The classical one-way ANOVA test requires an assumption of equal variances for all groups. In our example, the homogeneity of variance assumption turned out to be fine: the Levene test is not significant.

How do we save our ANOVA test, in a situation where the homogeneity of variance assumption is violated?

An alternative procedure (i.e.: Welch one-way test), that does not require that assumption have been implemented in the function `oneway.test()`.

ANOVA test with no assumption of equal variances

```
##
```

```
## One-way analysis of means (not assuming equal variances)
```

```
##
```

```
## data:  weight and group
```

```
## F = 5.181, num df = 2.000, denom df = 17.128, p-value =
```

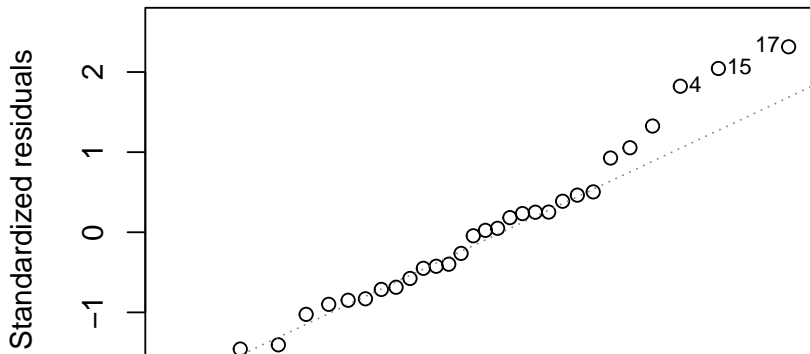
Pairwise t-tests with no assumption of equal variances

## Check the normality assumption

Normality plot of residuals. In the plot below, the quantiles of the residuals are plotted against the quantiles of the normal distribution. A 45-degree reference line is also plotted.

The normal probability plot of residuals is used to check the assumption that the residuals are normally distributed. It should approximately follow a straight line. As all the points fall approximately along this reference line, we can assume normality.

Normal Q-Q



# Kruskal-Wallis Test

Kruskal-Wallis test by rank is a non-parametric alternative to one-way ANOVA test, which extends the two-samples Wilcoxon test in the situation where there are more than two groups. It's recommended when the assumptions of one-way ANOVA test are not met. This tutorial describes how to compute Kruskal-Wallis test in R software.

## Compute Kruskal-Wallis test

```
##
```

```
## Kruskal-Wallis rank sum test
```

```
##
```

```
## data: weight by group
```

```
## Kruskal-Wallis chi-squared = 7.9882, df = 2, p-value = 0.01845
```

## Multiple pairwise-comparison between groups

From the output of the Kruskal-Wallis test, we know that there is a significant difference between groups, but we don't know which pairs of groups are different. It's possible to use the function `pairwise.wilcox.test()` to calculate pairwise comparisons between group levels with corrections for multiple testing.

```
## Warning in wilcox.test.default(xi, xj, paired = paired,
## exact p-value with ties

##
##  Pairwise comparisons using Wilcoxon rank sum test
##
## data:  PlantGrowth$weight and PlantGrowth$group
##
##      ctrl  trt1
## trt1 0.199 -
## trt2 0.095 0.027
##
## P value adjustment method: BH
```