Analysis of Variance

The *t-test* was used to test hypotheses about the means of two independent samples. For example, to test if there is deference between control and treatment groups. The method called analysis of variance (ANOVA) allows us to compare means for more than two independent samples.

The one way *anova* test is a hypothesis test to see if the means of the variables are all equal. Simply, it can be considered as a generalization of the two-sample *t-test*. To apply the analysis of variance data should be independent and normally distributed. Therefore before applying the *anova,* attention should be focused on the assumptions. Normal probability plot can be used to check the normality assumption.

The one-way test is a hypothesis test that tests the null hypothesis that $\mu_1 = \mu_2 = \cdots = \mu_p$ against that alternative that one or more means is deferent. That is

H_0: $\mu_1 = \mu_2 = \cdots = \mu_p$ H_1: at least one is not equal.

Two functions are useful in this example: ***oneway.test*** to perform the hypothesis test, and ***anova*** to give detailed.

Eg:-

Group A 10 12 14 23 25 26 27 27 31 33 36 37 38 40 42 43 64 65

Group B 8 12 22 24 24 26 30 33 35 35 38 39 41 44 45 50 52

Group C 22 24 26 35 38 39 39 40 43 45 46 53 55 57 60 62

>A <- c(10,12,14,23,25,26,27,27,31,33,36,37,38,40,42,43,64,65)

>B <- c(12,22,24,24,26,30,33,35,35,38,39,41,44,45,50,52)

>C <- c(22,24,26,35,38,39,39,40,43,45,46,53,55,57,60,62)

> scores <- data.frame(A,B,C)

> boxplot(scores) #boxplot allows us to compare the three distributions

Analysis of variance allows us to investigate if all the graders have the same mean. The R function to do the analysis of variance hypothesis test (oneway.test) requires the data to be in a deferent format. It wants to have the data with a single variable holding the scores, and a factor describing the grader or category. The stack command will do this for us:

> scores = stack(scores)

> names(scores) # scores can be seen

[1] "values" "ind"

To call *oneway.test* we need to use the model formula notation as follows

> oneway.test(values ~ ind, data=scores, var.equal=T)

#only do hypothesis test , var.equal=T set explicitly that the variances are equal

> anova(lm(values ~ ind, data= scores)) # gives more detail

The Kruskal-Wallis test

Normal distribution of data is a main assumption related to the anova. In the case of that assumption is in trouble non parametric test known as the *Kruskal-Wallis* can be used. It used in a similar manner as the *Wilcoxen signed-rank* test is used in place of the *t-test*. It too is a test on the ranks of the original data and so the normality of the data is not needed. The *Kruskal-Wallis* test will be appropriate if you don't believe the normality assumption of the *oneway test*. Its use in R is similar to *oneway.test*

> kruskal.test(values ~ ind, data= scores)