**Data Analytics Assignment 6\_4 – Session 6**

**Q1**

**a) What are the assumptions of ANOVA, test it out?**

**Solution:**

ANOVA is a statistical technique that assesses potential differences in a scale-level dependent variable by a nominal-level variable having 2 or more categories.

1. **Independence of cases** – this is an assumption of the model that simplifies the statistical analysis.

We need *r* simple random samples for the *r* treatments, and they need to be independent samples. The sample sizes need not be the same, though it’s best if they’re not very different.

1. **Normality** – the distributions of the residuals are normal

The underlying populations should be normally distributed. However, the ANOVA test is robust and moderate departures from normality aren’t a problem, especially if sample sizes are large and equal or nearly equal

1. **Equality (or "homogeneity") of variances**, called homoscedasticity

The samples should all have the same standard deviation, theoretically. Because the ANOVA test is robust, says it’s good enough if the largest standard deviation is less than double the smallest standard deviation.

When sample sizes are equal but standard deviations are not, the actual p-value will be slightly larger than what you find in the tables. But when sample sizes are unequal, and the smaller samples have the larger standard deviations, the actual p-value “can increase dramatically above” what the tables say, even “without too much disparity” in the standard deviations. “Falsely reporting significant results when the small samples have the larger variances is a serious worry.

**b) Why ANOVA test? Is there any other way to answer the above question?**

**Solution:**

Friedman's Test is a fairly standard non-parametric statistic that's analogous to a repeated measures ANOVA. If null hypothesis is rejected, friedmanmc() function does post hoc analysis.

**For example Please find R Script File in the repository:**

# Independence, normality & Equality

yeast <- read.table("E:/Data Analytics with RET/Assignment/yeast.txt", quote="\"", comment.char="")

names(yeast) <- c("seq","mcg", "gvh", "alm", "mit", "erl", "pox", "vac", "nuc", "class")

attach(yeast)

qqnorm(nuc)

qqline(nuc)

# Bartlett Test of Homogeneity of Variances

bartlett.test(nuc~class, data=yeast)

# Figner-Killeen Test of Homogeneity of Variances

fligner.test(nuc~class, data=yeast)

# Homogeneity of Variance Plot

library(HH)

hov(nuc~class, data=yeast)

hovPlot(nuc~class, data=yeast, str = 90, las = 3)