# One-Way Analysis of Variance

## 1 Introduction

Analysis of variance (ANOVA) is a widely used statistical technique. It is used in design of experiments, survey design, quality improvement, and many other statistical industries. The discussion to fully cover all the details relating to Analysis of Variance would take volumes of books. I have no intention to provide a comprehensive discussion on the topic. I will save a more intricate discussion on the subject for a later date. Here I will provide a basic discussion of a straight forward analysis.

When a researcher needs to compare the response for multiple levels Analysis of Variance is often the tool of choice. When there is only one factor with two levels then the analysis can be conducted using what is known as a *t-test*. However, when there are multiple factors or when there are three or more levels the t-test becomes less useful because a researcher runs into a multiple comparisons problem. The t-test is therefore considered a special case of the one-way ANOVA.

There are many different uses for ANOVA but here I'll describe the one-way ANOVA. This is one of the most basic uses for analysis of variance and will allow a researcher to identify differences that exist between groups where the response value is continuous (non-categorical).

## 2 Assumptions

- Observations are independent
- Errors are independent and identically distributed
- The errors have a normal distribution
- Homoscedastic variance – the variances of the groups should be the same

# 3 Example

Suppose a researcher is investigating the amount of time based on four different groups. The goal of the experiment is to determine if there is indeed a difference in any of the groups. For this hypothetical *completely random design* experiment 400 observation units were assigned to each of the four treatments.

For this example we test the homogeneity of variance and that the data is from a normal distribution. The below example uses the Shapiro-Wilk test for normality and uses the Bartlett Test of Homogeneity of Variances and the Fligner-Killeen Test of Homogeneity of Variances. The below code shows the approach when conducting a cell means model measuring against zero and an effects model and it also provides some basic summary statistics for the data.

```
set.seed(12345);
r <- data.frame( rbind(
  cbind(1,rnorm(100,5,1)),
  cbind(2,rnorm(100,6,1)),
  cbind(3,rnorm(100,8,1)),
  cbind(4,rnorm(100,10,1))
) );
names(r) <- c("group","value");
r$group <- as.factor(r$group);
fit  <-  aov(r$value ~ r$group);
fit.lm.effect <- lm(r$value ~ r$group);
fit.lm.cellmean<-lm(r$value ~ r$group-1);
summary(fit);
summary(fit.lm.effect);
summary(fit.lm.cellmean);
tapply(r$value, r$group, mean);
tapply(r$value, r$group, sd);


##Q-Q Plot
qqnorm(r$value); qqline(r$value, col=4);
##Shapiro-Wilk Normality Test
by(r$value, r$group, shapiro.test);


##Bartlett test of homogeneity of variance
bartlett.test(r$value ~ r$group);
##Fligner-Killeen test of homogeneity of variance
fligner.test(r$value ~ r$group);
```