1. Use the given link Data Set.

Answer the below questions:

1. What are the assumptions of ANOVA, test it out?

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

Normality – the distributions of the residuals are normal.

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

1. Why ANOVA test? Is there any other way to answer the above question?

Analysis of variance (ANOVA) is a collection of statistical models and their associated estimation procedures (such as the "variation" among and between groups) used to analyze the differences among group means in a sample. ANOVA was developed by statistician and evolutionary biologist Ronald Fisher. In the ANOVA setting, the observed variance in a particular variable is partitioned into components attributable to different sources of variation. In its simplest form, ANOVA provides a statistical test of whether the population means of several groups are equal, and therefore generalizes the t-test to more than two groups. ANOVA is useful for comparing (testing) three or more group means for statistical significance. It is conceptually similar to multiple two-sample t-tests, but is more conservative, resulting in fewer type I errors,[1] and is therefore suited to a wide range of practical problems.

One-way ANOVA

Sources df SS MSS F-ratio

Between treatment k−1 SST MST (SST/(k−1)) MST/MSE

Error N−k SSE MSE (SSE/(N−k))

Total N−1

SSE – Sum-of-Square due to Error

SST – Sum-of-Square of Treatment

MST – Mean Sum-of-square Treatment

MSE – Mean Sum-of-square Error

df – Degrees of Freedom

Two-way ANOVA

Source df SS MSS F-ratio

Between treatment k−1 SSR MST (SSR/(k−1)) MST/MSE (F-ratio row)

Between block h−1 SSC MSV (SSC/(h−1)) MSV/MSE (F-ratio column)

Error (h−1)(k−1) SSE MSE (SSE/((h−1)(k−1)))

Total N−1

SSR – Sum-of-Square of treatment in Rows (SST)

SSC – Sum-of-Square between Column

MSV – Mean Sum of Variance