One Way Anova: Concepts and Application in Agricultural System

Hussaini Abubakar[1], Haruna Danyaya Abubakar[2], and Aminu Salisu[3]
[1]Department of Mathematics and Statistics, Hussaini Adamu Federal Polytechnic kazaure, Jigawa, Nigeria
[2]Department of Science Laboratory Technology, Hussaini Adamu Federal Polytechnic kazaure, Jigawa, Nigeria
[3]Department of Science Laboratory Technology, Hussaini Adamu Federal Polytechnic kazaure, Jigawa, Nigeria
yankwanee85y@gmail.com

Abstract - Agriculturalists seek general explanations for the variations in agricultural yields in response to a treatment. An increasingly popular solution is the powerful statistical technique one way analysis of variance (ANOVA). This technique is intended to analyze the variability in data in order to infer the inequality among population means. After exploring the concept of the technique, the response of the chlorophyll content on the leaves of 160 maize seedlings to the treatment of nitrogen potassium phosphorous (NPK) at 0g, 5g, 10g and 20g as control treatment, treatment1, treatment2 and treatment3 respectively, it was revealed that, there was a significant effect of the amount of NPK on the chlorophyll content of maize seedlings at P < 0.05, [F (3, 141) = 51.190, P = 0.000]. Post hoc comparison using tukey HSD test indicated that, the mean score for treatment1 (M = 18.89, SD = 11.58) was significantly different than treatment3 (M =1.61, SD = 7.01) and the control treatment (M = 4.59, SD = 5.49), also the mean score for treatment2 (M =21.57, SD = 9.80) was significantly different than treatment3 and the control treatment respectively. However, the result indicated a non-significant difference between the treatments 1 and 2 and treatments 3 and the control treatment respectively. Altogether the result revealed that the amount of NPK really do have effect on the chlorophyll content of maize seedlings. The data were analyzed using computer program SPSS.

Keywords - one-way ANOVA test, multiple comparison tests, NPK, chlorophyll, SPSS.

1. Introduction

The concept of analysis of variance (ANOVA) was established by the British geneticist and statistician sir R. A. Fisher in 1918 and formally published in his book "statistical methods for workers" in 1925. The technique was developed to provide statistical procedures for test of significance for several group means. ANOVA can be conceptually viewed as an extension of the two independent samples t-test to multiple samples t-test, but results in less type 1 error and therefore suited a wide range of practical problems. Formerly, this idea was generally used for agricultural experiments, but is presently the most commonly advanced research method in business, economic, medical and social science disciplines.

Like many other parametric statistical techniques, ANOVA is based on the following statistical assumptions:

a) Homoscedasticity (homogeneity) of variance.
b) Normality of data.
c) Independence of observations.

2. Basic concepts of one way ANOVA test.

A one-way analysis of variance is used when the data are divided into groups according to only one factor.

Assume that the data $y_{11}, y_{12}, y_{13}, \ldots, y_{1n_1}$ are sample from population 1, $y_{21}, y_{22}, y_{23}, \ldots, y_{2n_2}$ are sample from population 2, $\cdots$, $y_{k1}, y_{k2}, y_{k3}, \ldots, y_{kn_k}$ are sample from population *k*. Let y_{ij} denote the data from the i^{th} group (level) and j^{th} observation.

We have values of independent normal random variables $Y_{ij} = 1, 2, 3, \ldots, k$ and $J = 1, 2, 3, \ldots, n_i$ with mean μ_i and constant standard deviation σ, $Y_{ij} \sim N(\mu_i, \sigma)$ Alternatively, each $Y_{ij} = \mu_i + \varepsilon_{ij}$ where ε_{ij} are normally distributed independent random errors, $\varepsilon_{ij} \sim N(0, \sigma)$. Let $N = n_1 + n_2 + n_3 + \ldots + n_k$ is the total number of observations (the total sample size across all groups), where n_i is sample size for the i^{th} group.

The parameters of this model are the population means μ_1, μ_2, μ_k and the common standard deviation σ.

Using many separate two-sample *t*-tests to compare many pairs of means is a bad idea because we don't get a *p*-value or a confidence level for the complete set of comparisons together.

We will be interested in testing the null hypothesis

$$H_0: \mu_1 = \mu_2 = \mu_k \tag{1}$$

against the alternative hypothesis

$$: \exists 1 \le i, l \le k: \mu_i \ne \mu_l \tag{2}$$

(there is at least one pair with unequal means).

Let $\bar{y}_i$ represent the mean sample *i* (*i = 1, 2, 3, …, k*):

$$\bar{y}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} y_{ij}\ , \tag{3}$$

$\overline{y}$ represent the grand mean, the mean of all the data points:

$$\overline{y} = \frac{1}{N}\sum_{i=1}^{k}\sum_{j=1}^{n_i} y_{ij}, \tag{4}$$

S_i^2 represent the sample variance:

$$S_i^2 = \frac{1}{n_i-1}\sum_{j=1}^{n_i}\left(y_{ij} - \overline{y}_i\right)^2, \tag{5}$$

and $S^2 = MSE$ is an estimate of the variance σ^2 common to all k populations,

$$S^2 = \frac{1}{N-k}\sum_{i=1}^{k}(n_i - 1).S_i^2. \tag{6}$$

ANOVA is centered around the idea to compare the variation between groups (levels) and the variation within samples by analyzing their variances.

Define the total sum of squares *SST*, sum of squares for error (or within groups) *SSE*, and the sum of squares for treatments (or between groups) *SSC*:

$$SST = \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(y_{ij} - \overline{y}\right)^2, \tag{7}$$

$$SSE = \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(y_{ij} - \overline{y}_i\right)^2 = \sum_{i=1}^{k}(n_i - 1).S_i^2, \tag{8}$$

$$SSC = \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(\overline{y}_i - \overline{y}\right)^2 = \sum_{i=1}^{k} n_i\left(\overline{y}_i - \overline{y}\right)^2, \tag{9}$$

Consider the deviation from an observation to the grand mean written in the following way:

$$y_{ij} - \overline{y} = \left(y_{ij} - \overline{y}_i\right) + \left(\overline{y}_i - \overline{y}\right). \tag{10}$$

Notice that the left side is at the heart of *SST*, and the right side has the analogous pieces of *SSE* and *SSC*. It actually works out that:

$$SST = SSE + SSC. \tag{11}$$

The total mean sum of squares *MST*, the mean sums of squares for error *MSE*, and the mean sums of squares for treatment *MSC* are:

$$MST = \frac{SST}{df(SST)} = \frac{SST}{N-1}, \tag{12}$$

$$MSE = \frac{SSE}{df(SSE)} = \frac{SSE}{N-k}, \tag{13}$$

$$MSC = \frac{SSC}{df(SSC)} = \frac{SSC}{k-1}, \tag{14}$$

The one-way ANOVA, assuming the test conditions are satisfied, uses the following test statistic:

$$F = \frac{MSC}{MSE}. \tag{15}$$

Under H_0 this statistic has Fisher's distribution $F(k-1, N-k)$. In case it holds for the test criteria

$$F > F_{1-\alpha,k-1,N-k}, \tag{16}$$

where $F_{1-\alpha,k-1,N-k}$ is $(1-\alpha)$quantile of F distribution with k - 1 and N - k degrees of freedom, then hypothesis H_0 is rejected on significance level α

The results of the computations that lead to the F statistic are presented in an ANOVA table, the form of which is shown in the table1.

Table1. Basic one way ANOVA table.

Source of Variation	*Sum Of Squares SS*	*Degrees of freedom df*	*Mean square*	*F - Statistic*	*Tail area above F*
Between group	*SSC*	*k-1*	*MSC*	*MSC/MSE*	*P – value*
Within	*SSE*	*N-k*	*MSE*		
Total	*SST*	*N-1*			

This *p*-value says the probability of rejecting the null hypothesis in case the null hypothesis holds. In case P < α, where α is chosen significance level, the null hypothesis is rejected with probability greater than (1-α) 100 probability.

3. Post hoc comparison procedures.

One possible approach to the multiple comparison problems is to make each comparison independently using a suitable statistical procedure. For example, a statistical hypothesis test could be used to compare each pair of means, μ_I and μ_J, $I, J = 1, 2, \ldots, k; I \neq J$, where the null and alternative hypotheses are of the form

$$H_0: \mu_I = \mu_J \text{ , } \mu_I \neq \mu_J \tag{17}$$

An alternative way to test for a difference between μ_I and μ_J is to calculate a confidence interval for $\mu_I - \mu_J$. A confidence interval is formed using a point estimate a margin of error, and the formula

$$\text{(Point estimate)} \pm \text{(Margin of error)}. \tag{18}$$

The point estimate is the best guess for the value of $\mu_I - \mu_J$ based on the sample data. The margin of error reflects the accuracy of the guess based on variability in the data. It also depends on a confidence coefficient, which is often denoted by 1-α. The interval is calculated by subtracting the margin of error from the point estimate to get the lower limit and adding the margin of error to the point estimate to get the upper limit.

If the confidence interval for $\mu_I - \mu_J$ does not contain zero (there by ruling out that ($\mu_I \neq \mu_J$), then the null hypothesis is rejected and μ_I and μ_J are declared different at level of significance α.

The multiple comparison tests for population means, as well as the *F*-test, have the same assumptions.

There are many different multiple comparison procedures that deal with these problems. Some of these procedures are as follows: Fisher's method, Tukey's method, Scheffé's method, Bonferroni's adjustment method, DunnŠidák method. Some require equal sample sizes, while some do not. The choice of a multiple comparison procedure used with an ANOVA will depend on the type of experimental design used and the comparisons of interest to the analyst.

The Fisher (LSD) method essentially does not correct for the type 1 error rate for multiple comparisons and is generally not recommended relative to other options.

The Tukey (HSD) method controls type 1 error very well and is generally considered an acceptable technique. There is also a modification of the test for situation where the number of subjects is unequal across cells called the Tukey-Kramer test.

The Scheffé test can be used for the family of all pairwise comparisons but will always give longer confidence intervals than the other tests. Scheffé's procedure is perhaps the most popular of the post hoc procedures, the most flexible, and the most conservative.

There are several different ways to control the experiment wise error rate. One of the easiest ways to control experiment wise error rate is use the Bonferroni correction. If we plan on making m comparisons or conducting m significance tests the Bonferroni correction is to simply use $^{\alpha}/_{m}$ as our significance level rather than α. This simple correction guarantees that our experiment wise error rate will be no larger than α. Notice that these results are more conservative than with no adjustment. The Bonferroni is probably the most commonly used post hoc test, because it is highly flexible, very simple to compute, and can be used with any type of statistical test (e.g., correlations), not just post hoc tests with ANOVA.

The Šidák method has a bit more power than the Bonferroni method. So from a purely conceptual point of view, the Šidák method is always preferred.

The confidence interval for $\mu_I - \mu_J$ is calculated using the formula:

$$\overline{Y}_I - \overline{Y}_J \pm t_{1-\alpha/2}\,, N-k\,.\sqrt{S^2\left(\frac{1}{n_I}+\frac{1}{n_J}\right)} \qquad (19)$$

where $t_{1-\alpha/2}\,, N-k$ is the quantile of the Student's t probability distribution, by Fisher method (LSD − Least Significant Difference);

$$\overline{Y}_I - \overline{Y}_J \pm q_{\alpha,k},\ N-k\,.\sqrt{\frac{S^2}{2}\left(\frac{1}{n_I}+\frac{1}{n_J}\right)}, \qquad (20)$$

where $q_{\alpha,k},\ N-k$ represents the quantile for the Studentized range probability distribution, by Tukey Kramer method (HSD − Honestly Significant Difference);

$$\overline{Y}_I - \overline{Y}_J \pm \sqrt{(k-1)S^2\left(\frac{1}{n_I}+\frac{1}{n_J}\right).F_{1-\alpha,k-1,N-k}} \qquad (21)$$

By Scheffé method;

$$\overline{Y}_I - \overline{Y}_J \pm t_{1-\frac{\alpha^*}{2}},\ N-k\,\sqrt{S^2\left(\frac{1}{n_I}+\frac{1}{n_J}\right)} \qquad (22)$$

where $\alpha^* = \frac{\alpha}{2}$, C = $\binom{K}{2}$ is the number of pairwise comparisons in the family, by Bonferonni method;

$$\overline{Y}_I - \overline{Y}_J \pm t_{1-\frac{\alpha^*}{2}},\ N-k\,\sqrt{S^2\left(\frac{1}{n_I}+\frac{1}{n_J}\right)} \qquad (23)$$

where $\alpha^* = 1-(1-\alpha)^{1/C}$ and $C = \binom{K}{2}$, by DunnŠidák metho

Test for homogeneity of variance

Many statistical procedures, including analysis of variance, assume that the different populations have the same variance. The test for equality of variances is used to determine if the assumption of equal variances is valid.

We will be interested in testing the null hypothesis

$$H_0: \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_k^2 \qquad (24)$$

against the alternative hypothesis

$$H_1: \exists 1 \le i, l \le k: \sigma_i^2 \ne \sigma_l^2 \qquad (25)$$

There are many tests assumptions of homogeneity of variances. Commonly used tests are the Bartlett (1937), Hartley (1940, 1950), Cochran (1941), Levene (1960), and Brown and Forsythe (1974) tests. The Bartlett, Hartley and Cochran are technically test of homogeneity. The Levene and Brown and Forsythe methods actually transform the data and then tests for equality of means.

Note that Cochran's and Hartley's test assumes that there are equal numbers of participants in each group.

The tests of Bartlett, Cochran, Hartley and Levene may be applied for number of samples k > 2. In such situation, the power of these tests turns out to be different. When the assumption of the normal distribution holds for k > 2 these tests may be ranked by power decrease as follows: Cochran Bartlett Hartley Levene. This preference order also holds in case when the normality assumption is disturbed. An exception concerns the situations when samples belong to some distributions which have more heavy tails then the normal law. For example, in case of belonging samples to the Laplace distribution the Levene test turns out to be slightly more powerful than three others.

Bartlett's test has the following test statistic:

$$B = C^{-1}\left[(N-k).InS^2 - \textstyle\sum_{i=1}^{k}(n_i-1).InS_i^2\right], \quad (26)$$

Where constant $C = 1 + \frac{1}{3(k-)}.\left(\sum_{i=1}^{k}\frac{1}{n_i-1} - \frac{1}{N-k}\right)$ and meaning of all other symbols is evident (see section 2). The hypothesis H_0 is rejected on significance level α, when

$$B > X^2{}_{1-\alpha,k-1} \qquad (27)$$

where $X^2{}_{1-\alpha,k-1}$ is the critical value of the *chi-square* distribution with k - 1 degrees of freedom.

Cochran's test is one of the best methods for detecting cases where the variance of one of the groups is much larger than that of the other groups. This test uses the following test statistic:

$$C = \frac{max s_i^2}{\sum_{i=1}^{k} S_i^2} \qquad (28)$$

The hypothesis H_0 is rejected on significance level α, when

$$C > C_{\alpha,k,n-1} \qquad (29)$$

where critical value $C_{\alpha,k,n-1}$ is in special statistical tables.

Hartley's test uses the following test statistic:

$$H = \frac{max S_i^2}{min S_i^2}. \qquad (30)$$

The hypothesis H_0 is rejected on significance level α, when

$$H > H_{\alpha,k,n-1}, \qquad (31)$$

where critical value $H_{\alpha,k,n-1}$ is in special statistical tables

Originally Levene's test was defined as the one-way analysis of variance on $Z_{ij} = |y_{ij} - \overline{y}_i|$, the absolute residuals $y_{ij} - \overline{y}_i$, I = 1, 2, 3, …, k and j = 1, 2, 3, …, n_i where

k is the number of groups and n_i the sample size of the i^{th} group. The test statistic has Fisher's distribution $F(k-1, N-k)$ and is given by:

$$F = \frac{(N-k)\sum_{i=1}^{k} n_i \cdot (Z_i - \overline{Z})^2}{(k-1)\sum_{i=1}^{k}\sum_{j=1}^{n_i}(Z_{ij} - \overline{Z}_i)^2}. \quad (32)$$

Where $N = \sum_{i=1}^{k} n_i$, $\overline{Z}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} Z_{ij}$, $\overline{Z} = \frac{1}{N}\sum_{i=1}^{k}\sum_{j=1}^{n_i} Z_{ij}$

To apply the ANOVA test, several assumptions must be verified, including normal populations, homoscedasticity, and independent observations. The absolute residuals do not meet any of these assumptions, so Levene's test is an approximate test of homoscedasticity.

Brown and Forsythe subsequently proposed the absolute deviations from the median $\tilde{y}_i$ of the i^{th} group, so is $Z_{ij} = |y_{ij} - \tilde{y}_i|$.

4. Methodology

The study was undertaken in Kazaure north jigawa Nigeria. The population for this study was one hundred and sixty (160) maize seedlings grown and studied for three weeks period. Information was collected from the target population (maize seedlings) with the aid of chlorophyll meter (SPAD 502 plus) to measure the chlorophyll content of the leaves of each seedling. Data analysis was with the aid of inferential statistics (one way ANOVA). Independent variable for the study was the amount of NKP measured in gram. The significance test for the between treatment effect was the researcher's statistical evidence of the effect of the treatment on the chlorophyll content of the leaves of the maize seedlings.

4.1. Test for normality and homogeneity of the data.

To begin ANOVA test, one must verify the validity of the normality and homogeneity assumptions of the data under study. These tests were based on Kolmongorov – Siminov and levene's statistic respectively. These normality and homogeneity tests were conducted and found tenable P > 0.05, at 0.05 level for all the four treatment levels and P < 0.05, at 0.05 level respectively. The results were presented in tables 2 and 3 below.

Table 2. Kolmongorov–Siminov test of normality.

TREATMENT	Statistic	df	Sig.
TREATMENT1	0.117	27	0.200*
TREATMENT2	0.103	43	0.200*
TREATMENT3	0.539	39	0.120*
CONTROL	0.298	36	0.130*

*. This is a lower bound of the true significance.

Table 3. Levene's test for homogeneity of variance.

Levene Statistic	df1	df2	Sig
11.159	3	141	0.000

4.2. Test of significance for the treatment effect.

After the tests for the assumption of normality and equality of variance (Homoscedesticity), the next thing is to determine the significant effect of the independent variable, in this case amount of nitrogen. The significance of the treatment is based on F distribution, the test revealed that the probability of the Fisher distribution F (3, 141) was 0.000, less than the level of significance of 0.05 (i.e, P < 0.05). The null hypothesis that there was no significant difference between the mean chlorophyll was rejected. As presented in table 2.

Table 4. One way ANOVA table for the experiment data.

Source of Variation	Sum Of Squares SS	df	Mean square	F
Between group	11373.06	3	3791.02	51.19
Within group	10442.14	141	74.058	
Total	21815.2	144		

4.3. Post hoc comparison.

When the null hypothesis is rejected using the *F*-test in ANOVA, we want to know where the difference among the means is. To determine which pairs of means are significantly different, and which are not, we can use the multiple comparison tests, in this case, tukey HSD. The result was presented in table 5.

Table 5. Post hoc comparison.

Pairs I, J	Mean Difference	Lower Bound	Upper Bound
C, 1*	-14.30648	-20.0027	-8.6108
C, 2*	-16.98366	-22.0381	-11.9292
C, 3	2.97842	-2.1928	8.1497
1, 2	-2.67717	-8.1711	2.8167
1, 3*	17.2849	11.6834	22.8864
2, 3*	19.96208	15.0146	24.9096

*The mean difference is significant at the 0.05 level.

7. Conclusion

In many statistical applications in agriculture, business administration, psychology, social science, and the natural sciences we need to compare more than two groups. For hypothesis testing more than two population means, scientists have developed ANOVA method. The ANOVA test procedure compares the variation in observations between samples (sum of squares for groups, *SSC*) to the variation within samples (sum of squares for error, *SSE*). The ANOVA *F* test rejects the null hypothesis that the mean responses are not equal in all groups if *SSC* is large relative to *SSE*. The analysis of variance assumes that the observations are normally and independently distributed with the same variance for each treatment or factor level.

However, the ANOVA F test revealed a significant effect of the amount of NPK on the chlorophyll content of maize seedlings at $P < 0.05$, [$F (3, 141) = 51.190$, $P = 0.000$], and also the tukey HSD test result indicated a non-significant difference between the treatments 1 and 2 and treatments 3 and the control treatment respectively. Altogether the results revealed that the amount of NPK really do have effect on the chlorophyll content of maize seedlings.

References

1. Aczel, A.D., *Comple Business Statistics*, (Irwing, 1989)
2. Brown, M., Forsythe, A., “Robust tests for the equality of variances,” journal of the American Statistical Association, **365-367** (1974)
3. Montgomery, D.C., Runger, G.C., *Applied Statistics and Probability for Engineers*, (John wiley & Sons, 2003)
4. Ostertagova, E., *Applied Statistics* (in Slovac), Elfa, Kosice, 2011.
5. Parra-Frutos, I., “ The bahaviour of the modified levene’s test when data are not normally distributed,” comput Stat, Springer, **671-693** (2009)
6. Rafter, J.A., Abell, M.L., Braselton. J.P., “Multiple Comparison Methods for Means,” SIAM Review, **44(2). 259-278** (2002)
7. Rykov, V.V., Balakrisnan, N., Nikulin, M.S., *Mathematical and Statistical Models and methods in Relability, Springer*, (2010).
8. Stephens, L.J., *Advanced Statistics demystified*, McGraw-Hill (2004)
9. Aylor, S., Business Statistics.www.palgrave.com.