Mathematical Theory and Modeling www.iiste.org
ISSN 2224-5804 (Paper) ISSN 2225-0522 (Online)
Vol.5, No.13, 2015

Determination of Single Factor Fixed Effect Experiments in CRD with Multiple Linear Regression Analysis

Hanaw A. Amin
Faculty of Science & Science Education, School of Science, University of Sulaimani,
Mathematics Department, Kurdistan, Iraq
hanaw.ammin@univsul.edu.iq or hanaw_a@yahoo.com

Abstract:
Regression analysis and ANOVA (Analysis of Variance) are two methods in the statistical theory to analyze the behavior of one variable compared to another. Usually statisticians deal with regression models and analysis of variance models as separate subjects, especially when refined in the initial levels. In fact, can be deal with analysis of variance model as a special case of the multiple regression models. Therefore, any model of analysis of variance can be solved by using multiple regression analysis and access to same results of matching to analysis of variance. The regression analysis is widely used in predicting and forecasting. It is also used to establish relationships in experimental data. Dummy variables are variables that take the values of only 0 or 1. They may be explanatory or outcome variables; however, the focus of this article is explanatory or independent variable construction and usage. We do not recommend using multiple regression method instead of the analysis of variance method, but it can be shown that however the multiple regression analysis method contain more than one method of analysis operations with respect to analysis of variance, but the use is more useful than in some designs especially in experimental data. Interpretation of the relationship between the analysis of variance of single factor fixed effect balanced CRD where all eplications n_i are equal and multiple regression analysis model by using dummy variables, and a more detailed introduction of a teaching techniques is presented.

Keywords: Fixed Model, Analysis of Variance, Complete Randomized Design, Dummy Variables, Multiple Regression Analysis.

1. Introduction:

In statisics regression analysis includes any technique for modeling and analyzing trends between a dependent variable and an independent variable. Regression analysis was first developed by Sir Francis Galton in the latter part of the19th century. Galton had studied the relation between heights of parents and children and noted that the heights of children of both tall and short parents appeared to "revert" or "regress" to the mean of the group. He considered this tendency to be a regression to "mediocrity." Galton developed a mathematical description of this regression tendency, the precursor of today's regression models. The term regression persists to this day to describe statistical relations between variables.

Regression analysis helps us to make predictionas outside of the given data. If the prediction is within the range of the given data it is called interpolation. If the prediction is outside the range of given data values it is called extrapolation (Toutenburg 2009).

Synonyms for dummy variables are design variables [Hosmer and Lemeshow, 1989] , Boolean indicators, and proxies [Kennedy, 1981]. Related concepts are binning [Tukey, 1977] or ranking, because belonging to a bin or rank could be formulated into a dummy variable. Bins or ranks can also function as sets and dummy variables can represent non-probabilistic set membership. Set theory is usually explained in texts on computer science or symbolic logic. Dummy variables are variables that take the values of only 0 or 1. They may be explanatory or outcome variables (Garavaglia and Sharma 1998). Typically, dummy variables are used in the following applications: time series analysis with seasonality or regime switching; analysis of qualitative data, such as survey responses; categorical representation, and representation of value levels. In a regression model, a dummy variable with a value of 0 will cause its coefficient to disappear from the equation. Conversely, the value of 1 causes the coefficient to function as a supplemental intercept, because of the identity property of multiplication by 1 (Garavaglia and Sharma1998). This type of specification in a linear regression model is useful to define subsets of observations that have different intercepts and/or slopes without the creation of separate models. In addition to the direct benefits to statistical analysis, representing information in the form of dummy variables is makes it easier to turn the model into a decision too (Seltman 2014).

The analysis of variance, which was originally developed by R.A. Fisher for field experiments, is one of the most widely used and one of the most general statistical procedures for testing and analyzing data. The corresponding F- test is a generalization of the t- test that compares two normal distributions. In general, this comparison is called comparison of the effects of treatments. If specific treatments are to be compared, then it is wise not to choose them at random, but to assume them as fixed (Neter, Kutner, Wasserman, and Nachtsheim 1996).

2. Multiple Regression Analysis:

The general purpose of multiple regression (the term was first used by Pearson, 1908) is to learn more about the relationship between several independent or predictor variables and a dependent or criterion variable. Therefore, the regression analysis is widely used in predicting and forecasting. It is also used to establish relationships in experimental data, in the fields of physics, chemistry, and many natural sciences and engineering disciplines. If the relationship or the regression function is a linear function, then the process is known as a linear regression. In the scatter plot, it can be represented as a straight line. If the function is not a linear combination of the parameters, then the regression is non-linear (Rawlings, Pantula, and Dickey 1998). Impose that we are dealing with two models, the first multiple regression model (Berger, and Kee 2009):

$$Y_j = \beta_0 + \beta_1 X_{1j} + \beta_2 X_{2j} + \dots + \beta_{kj} + \epsilon_j \quad (1)$$

And completely randomized fixed effect design:

$$Y_{ij} = \mu + t_i + \in_{ij} = \mu_i + \epsilon_{ij} \quad (2)$$

In multiple linear regression model expresses the dependent variable in terms of one or more independent variables and the goal is to be a forecasting of the variable dependent variable through the information contained in the independent variables. Independent variables may be correlated with each other and also with the dependent variable. Multiple regression is a direct extension of simple regression to multiple explanatory variables. Each new explanatory variable adds one term to the structural model.

The fixed model effects are used for the multiple comparisons of means of quantitative normally distributed factors that are observed on fixed selected experimental units. The term assumption in statistics refers to any part of a statistical model. For one-way ANOVA, the assumptions are normality, equal variance, and independence of errors. Correct assignment of individuals to groups is sometimes considered to be an implicit assumption. The statistical model for which one-way ANOVA is appropriate is that the (quantitative) outcomes for each group are normally distributed with a common variance (σ^2). The errors (deviations of individual outcomes from the population group means) are assumed to be independent. The model places no restrictions on the population group means (Milliken and Johnson 2009).

The null hypothesis test $H_0: \mu_1 = \mu_2 = \cdots = \mu_k$ against the general alternative H_1: at least two means are different, comparing (k) normally distributed populations with respect to their means. Since these quantitative variables are usually independent and not correlated this makes calculations easier in analysis of variance comparing to the regression analysis (Toutenburg 2009),. Method of matrix will be used to describe the features of the two models (1)and (2). The regression model (Muhammad 2004):

$$\underline{Y} = X\underline{\beta} + \underline{\in} \quad (3)$$

In detail

$$\begin{bmatrix} Y_1 \\ Y_2 \\ . \\ . \\ . \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & X_{21} & \cdots & X_{k1} \\ 1 & X_{21} & X_{22} & \cdots & X_{k2} \\ . & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ 1 & X_{n1} & X_{n2} & \cdots & X_{kn} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ . \\ . \\ . \\ \beta_k \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ . \\ . \\ . \\ \varepsilon_n \end{bmatrix}$$

Where the vector on the right of the equal represents responses in the experiment. A matrix (X) is a matrix of predictor variables and vector ($\underline{\beta}$) is a vector of parameters that appear in the model. The vector $\underline{\in}$ is a random error. The parameters in the vector ($\underline{\beta}$) can be estimated using the method of least squares from the equation (Muhammad 2004), (Kleinbaum, Kupper, Nizam 2008):

$$(X'X)\hat{\beta} = X'Y$$

Where $(X'Y)$ must be non- singular matrix, and $(X'Y)$ is the vector sum of multiplication elements in columns (X) and elements of the vector $(\underline{Y})$, and $(\underline{\hat{\beta}})$ is a vector of estimates of parameters that appear in the model, and it estimates can be obtained as follows (Muhammad 2004):

$$(X'X)^{-1}XY = \underline{\hat{\beta}} \qquad (4)$$

The analysis of variance in matrix model (Kleinbaum, Kuppere, Nizamr, Rosenberg2008), (Berger, and kee 2009), (Rawlings, Pantula, and Dickey 1998): :

$$\begin{bmatrix} Y_{11} \\ . \\ . \\ . \\ Y_{1n} \\ . \\ . \\ . \\ Y_{kn} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & \dots & 0 \\ . & . & & & . \\ . & . & & \dots & . \\ . & . & & & . \\ 1 & 1 & 0 & \dots & 0 \\ 1 & 0 & 1 & \dots & 0 \\ . & . & & & \\ . & . & 1 & \dots & 0 \\ . & . & & & \\ 1 & 0 & 0 & \dots & 1 \end{bmatrix} \begin{bmatrix} \mu \\ t_1 \\ t_2 \\ . \\ . \\ . \\ t_k \end{bmatrix} + \begin{bmatrix} \mu \\ \varepsilon_{11} \\ \varepsilon_{12} \\ . \\ . \\ . \\ \varepsilon_{kn} \end{bmatrix}$$

Where the matrix (X) consists of zero or one values, or is defined as (Toutenburg 2009),(Gary 2010), (Neter 1996) , (Muhammad 2004):

$$\boldsymbol{X}_i = \begin{cases} 1, & \textit{if the observed value of } Y_{ij} \textit{ belong to } i^{th} \textit{ treatment} \\ 0, & \textit{if the observed value of } Y_{ij} \textit{ belong to other treatment} \end{cases}$$

Where $i = 1, 2, \dots, k$ number of kth treatment which is represented as a dummy variable. A dummy variable or indicator variable is an artificial variable created to represent an attribute with two or more distinct categories/levels (Skrivanek2009). Dummy variables play an important role in the analysis of data, whether they are real-valued variables, categorical data, or analog signals. The extreme case of representing all the variables (independent and dependent) as dummy variables provides a high degree of flexibility in selecting a modeling methodology. In addition to this benefit of flexibility, the elementary statistics (e. g., mean and standard deviation) for dummy variables have interpretations for probabilistic reasoning, information theory, set relations, and symbolic logic. Whether the analytical technique is traditional or experimental, highly complex information structures can be represented by dummy variables. The intelligent use of dummy variables usually makes the resulting application easier to implement, use, and interpret. Dummy variables based on set membership can help when there are too few observations, and thus, degrees of freedom, to have a dummy variable for every category or some categories are too rare to be statistically significant (Garavaglia and Sharma 1998).

The least square estimated in the analysis of variance is:

$$\begin{bmatrix} nk & n & n & \dots & n \\ n & n & 0 & \dots & 0 \\ n & 0 & n & \dots & 0 \\ . & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ n & 0 & 0 & \dots & n \end{bmatrix} \begin{bmatrix} \hat{\mu} \\ \hat{t}_1 \\ \hat{t}_2 \\ . \\ . \\ . \\ \hat{t}_k \end{bmatrix} = \begin{bmatrix} Y.. \\ Y_1. \\ Y_2. \\ . \\ . \\ . \\ Y_k. \end{bmatrix} \qquad (5)$$

In other word $(X'X)\hat{\beta} = X'Y$, it is easy to explain why the difference between these two models. The last columns, with (k^{th}) number in the matrix $(X'X)$ in the analysis of variance model is equal to the first column and the matrix $(X'X)$has no inverse and this means there is no single solution for estimating parameters. The statistical tests $H_0: t_i = 0$ to the mean differences in the analysis of variance under the condition$\sum_{i=1}^{k} t_i = 0$ are (Muhammad 2004), (Seltman 2014):

$$\mu_1 = \mu + t_1, \quad \mu_2 = \mu + t_2, \dots, \quad \mu_k = \mu + t_k,$$

Under the condition $\sum \hat{t}_i = 0$ in the case of equal replications for estimating the parameters using least square methods are:

$$\hat{\mu} = \frac{Y..}{nk} = \bar{Y}.., \quad \hat{t}_i = \frac{Y_i.}{n} - \frac{Y..}{nk} = \bar{Y}_i. - \bar{Y}.., \quad i = 1,..,k$$

But in the case of unequal replications the hypothesis tests $H_0: t_i = 0$ to the mean differences using least square methods under the condition$\sum_i n_i t_i = 0$ the estimated parameters are (Berger, and Kee 2009), (Kleinbaum, Kupper, Nizam, Rosenberg 2008):

$$\hat{\mu} = \frac{Y_{..}}{N} = \bar{Y}.. \quad \text{and} \quad \hat{t}_i = \frac{Y_i}{n_i} - \frac{Y_{..}}{N}$$

For testing the hypotheses $H_0: t_i = 0$ using the technique of multiple regression and the sum of square of regression parameters$t_1, t_2, \dots, t_k$, which is equivalent to $\beta_1, \beta_2, \dots, \beta_k$, the sum of squares due to (uncorrected) regression isRawlings, Pantula, and Dickey 1998), (Muhammad 2004):

$$R(\mu, t_1, t_2, \dots, t_k) = \hat{\mu}Y.. + \hat{t}_1 Y_{1.} + \hat{t}_2 Y_{2.} + \dots + \hat{t}_k Y_{k.} \qquad (6)$$

$$= \bar{Y}..Y.. + (\bar{Y}_{1.} - \bar{Y}..)Y_{1.} + (\bar{Y}_{2.} - \bar{Y}_{..})Y_{2.} + \dots + (\bar{Y}_{k.} - \bar{Y}_{..})Y_{k.}$$

$$= \left(\frac{Y_{..}^2}{nk} + \frac{\sum Y_{i.}^2}{n}\right) - \bar{Y}_{..}(\sum_{i=1}^{k} Y_{i.})$$

$$= \frac{\sum Y_{i.}^2}{n} + \frac{Y_{..}^2}{nk} - \frac{Y_{..}^2}{nk}$$

$$= \frac{\sum Y_{i.}^2}{n} = \hat{\beta}' X' \underline{Y}$$

In other words, the relationship in equation (6) represents the sum of the product of the $\hat{\mu}, \hat{t}_1, \dots, \hat{t}_k$ within the right side of the equation (5). Assuming that the alternative is true, it means that the mean of $t_i = 0$ using the least squares method, the normal equations in (5) will be reduced as below (Berger and Kee 2009), Rawlings ,Pantula, and Dickey 1998), (Mason, Gunst , Hess 2003), (Vuchkov and Boyadjieva 2009), (Muhammad 2004):

$$\hat{\mu} = \frac{Y_{..}}{nk} \quad \text{or} \quad nk\hat{\mu} = Y_{..} \tag{7}$$

$$R(\mu) = \left(\frac{Y_{..}}{nk}\right)(Y_{..}) = \frac{Y_{..}^2}{nk}$$

Multiplying the right hand of equation (5) by $\hat{\mu}$, then the sum square due to $\hat{t}_1, \dots, \hat{t}_k$ if by μ is known or the sum of square due to (corrected) regression is (Muhammad 2004):

$$R(t_1, t_2, \dots, t_k | \mu) = R(\mu, t_1, t_2, \dots, t_k) - R(\mu)$$

$$= \frac{\sum Y_{i.}^2}{n} - \frac{Y_{..}^2}{nk} = SSTr.$$

With (k-1) degrees of freedom, then the sum square due to residual is:

$$SSE = \left(\sum_{i=1}^{k}\sum_{j=1}^{n} Y_{ij}^2 - \frac{Y_{..}^2}{nk}\right) - R(t_1, t_2, \dots, |\mu)$$

$$SSE = \sum_{i=1}^{k}\sum_{j=1}^{n} Y_{ij}^2 - \frac{Y_{..}^2}{nk} - \frac{\sum_{i=1}^{k} Y_{i.}^2}{n} + \frac{Y_{..}^2}{nk}$$

$$= \sum_{i=1}^{k}\sum_{j=1}^{n} Y_{ij}^2 - \frac{\sum_{i=1}^{k} Y_{i.}^2}{n}$$

With (nk-1)-(k-1)=k(n-1) degrees of freedom in the regression analysis to test the hypothesis $H_0: t_i = 0$ at $\alpha = 0.05$ level of significance, and the calculated valued of F will be computed from the formula of(Muhammad 2004) :

$$F = \frac{R(t_1, t_2, \dots, t_k|\mu)/(k-1)}{SSE/[k(n-1)]}$$

$$= \frac{SSTr./(k-1)}{SSE/[k(n-1)]}$$

When the calculated valued of F is greater or equal than the tabulated value $F_\alpha[(k-1), k(n-1)]$ from F- distribution, the null hypothesis is rejected (Toutenburg, 2009), (Rawlings, Pantula, and Dickey 1998),

3- Practical Part:

Teachers who attempt to use inquiry-based, student-centered instructional tasks face challenges that go beyond identifying well-designed tasks and setting them up appropriately in the classroom. Because solution paths are usually not specified for these kinds of tasks, students tend to approach them in unique and sometimes unanticipated ways. Teachers must not only strive to understand how students are making sense of the task but also begin to align students' disparate ideas and approaches with canonical understandings about the nature of mathematics.

In this article, we present a three scientific model that specifies key practices teachers can learn to use student responses to such tasks more effectively in discussions: classical technique, data show and classical technique with data shows technique. The goal is to compare between multiple regression and analysis of variance method through three different teaching methods, at mathematics department, which is characterized by theoretical nature. Examinations records of thirty participants were randomly chosen from a population from each three different teaching techniques as shown in Table (1).

Table .1Examination Results of Three Teaching Techniques

Data Show	Classic with Data Show	Classical Technique
12.0	11.0	45.0
11.0	25.0	27.0
17.0	23.0	31.0
26.0	21.0	27.0
20.0	35.0	33.0
31.0	18.0	26.0
31.0	37.0	32.0
13.0	29.0	29.0

I- ANOVA in practices :

1- The experimental model:

$$Y_{ij} = \mu + t_i + \varepsilon_{ij}$$

which (μ) is the overall mean, t_i is the effect of the i^{th} level of treatment (i.e., the deviation treatment effect from the overall mean(μ) caused by the i^{th} level) where $(i = 1, 2, 3), (j = 1, 2, \ldots, 8)$ and ε_{ij} is a random error and $\sum_{i=1}^{k} t_i = 0$, where $(i = 1, 2, 3)$.

2- The Hypotheses Test

$H_0: t_i = 0$

$H_1: t_i \neq 0$

3- These procedures require a large amount of computation, especially in the case of complicated classifications. For this reason, these procedures are available as software like SPSS. The ANOVA Table results are from SPSS program package as shown in Table (2):

Table . 2 ANOVA of Completely Randomizes Design

S .O . V.	Sum of Squares	Df	Mean Square	F	Sig.	$F_{\propto}(df_1, df_2)$
Between Groups	498.583	2	249.292	4.131	.031	$F_{(0.05,2,23)} = 3.42$
Within Groups	1267.250	21	60.345			
Total	1765.833	23				

Tabulated value is $F_{(0.05,2,23)} = 3.42$to be compared with calculated F- valued. Since the calculated F = 4.131 is greater than the tabulated values (3.42), with a significant p-value = 0.031 the null hypothesis of equally of treatment effects is rejected. It may conclude that at ($\propto = 0.05$) levels of significance at least one pair of population technique means are not equal (DeCster 2006). The data do provide evidence to indicate a difference in teaching techniques. It can be conclude that the best technique is classical method as shown in Fig. (1) and Fig. (2) of different means value,

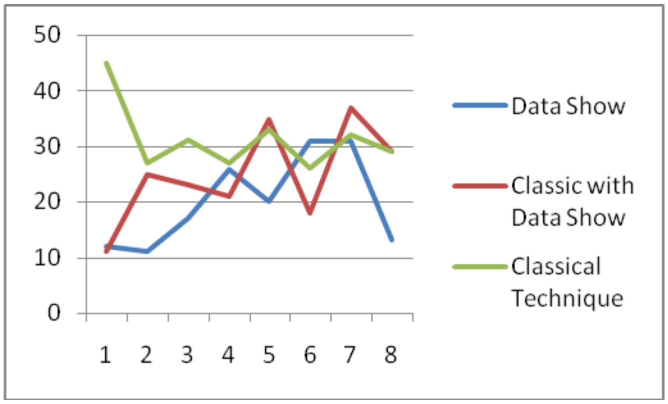

Figure1. Individual's Scores Line

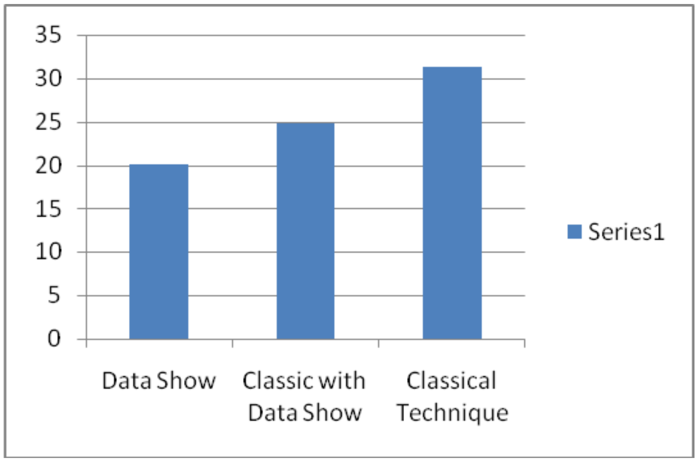

Figure 2.Means of Treatment Groups

LSD test for mean differences between techniques mean gives a significance differences level between data show and classical techniques as $|\bar{Y}1. - \bar{Y}3.| = 11.124 > t_{0.025}(21)\sqrt{15.086} = 8.079$ (DeCster 2006), at p- valued equal to (0.009), that means the null hypothesis $H_0 = \mu_i - \mu'_i = 0$ is rejected. It may conclude that there is a strong statistically difference between population data show and classical technique means as shown in Table (3) (Seltman 2014).

Table. 3 Three Multiple Comparisons of Least- Significant Difference (LSD) Test

(I) group	(J) group	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
1	2	-4.75	3.88411	0.235	-12.8274	3.3274
	3	-11.12500*	3.88411	0.009	-19.2024	-3.0476
2	1	4.75	3.88411	0.235	-3.3274	12.8274
	3	-6.375	3.88411	0.116	-14.4524	1.7024
3	1	11.12500*	3.88411	0.009	3.0476	19.2024
	2	6.375	3.88411	0.116	-1.7024	14.4524

*. The mean difference is significant at the 0.05 level.

II Linear Regression Analysis in Practices:

Multiple regressions is a straightforward extension of simple regression from one to several quantitative explanatory variables (and also categorical variables).

1- The Regression Model:

From the decline in the previous section the model of regression led to the matrix has no inverse, and to overcome this problem in the application, the following regression model is used:

$$Y_{ij} = \beta_0 + \beta_1 x_{1j} + \beta_2 x_{2j} + \epsilon_{ij}$$

Where $i = 1, 2, 3$ and $j = 1, 2, 3, 4$ and;

1- The Hypotheses Test:

$H_0: \beta_i = 0$

$H_1: \beta_i \neq 0$

Categorical variables in multiple regression To use a categorical variable with (k) levels in multiple regression we must re-code the data column as (k-1) new columns, each with only two different codes (most commonly we use 0 and 1). Variables that only take on the values 0 or 1 are called indicator or dummy variables. They should be considered as quantitative variables and should be named to correspond to their “1" level (Seltman 2014).

An indicator variable is coded 0 for any case that does not match the variable name and 1 for any case that does match the variable name (DeCster 2006).

$$x_i = \begin{cases} 1, & if\ Y_{ij}\ is\ observed\ to\ i^{th}\ treat \\ 0, & if\ Y_{ij}\ is\ observed\ to\ other\ treat \end{cases}$$

Where $i = 1,2$; and $j = 1, 2, \ldots, n$ and $(k-1)$variables are used to represent the treatment effects. It can clarify the relationship between the parameters of regression model $\beta_0, \beta_1, \beta_2$ and between the parameters of analysis of variance model $\mu, \tau_1, \tau_2, \tau_3$:

The linear regression model for the observed value of Y_{1j} when $x_{1j} = 1$ and $x_{2j} = 0$ is:

$$Y_{1j} = \beta_0 + \beta_1(1) + \beta_2(0) + \epsilon_{1j}$$

$$= \beta_0 + \beta_1 + \epsilon_{1j}$$

While the first observed dependent variable Y_{1j} in the regression model is:

$$Y_{1j} = \mu + \tau_1 + \epsilon_{1j}$$

$$= \mu_1 + \epsilon_{1j}$$

and so on

$$\mu_1 = \beta_0 + \beta_1$$

and in the same way it can prove

$$\mu_2 = \beta_0 + \beta_2$$

The linear regression model for the observed value for the third treatment Y_{3j} is:

$$Y_{3j} = \beta_0 + \beta_1(0) + \beta_2(0) + \epsilon_{ij}$$

$$= \beta_0 + \epsilon_{ij}$$

Whereas $x_{1j} = 0$, and $x_{2j} = 0$. In other word $\mu_3 = \beta_0$, and the relationship between the parameters of the two models is:

$$\mu_3 = \beta_0$$

$$\mu_1 - \mu_3 = \beta_1$$

$$\mu_2 - \mu_3 = \beta_2$$

In general if there is $(k)\,th$ treatments then the linear regression model consist of $(k-1)$ dummy variables

$$Y_{ij} = \beta_0 + \beta_1 x_{ij} + \beta_{2j} + \dots + \beta_{k-1} x_{k-1,j} + \epsilon_{ij}$$

The relation between the parameters of two models is:

$$\beta_0 = \mu_k$$

$$\beta_i = \mu_i - \mu_k,$$

Where $i = 1, 2, \dots, k-1$. $\beta_0$is always represent the estimate of the average of last treatment effects, and $\beta_i$is the estimate of the difference between two average $(i \text{ and } k)$ treatment effects. In other words:

$$\mu_k = \beta_0$$

$$\mu_i - \mu_k = \beta_i, \quad i = 1, 2, \dots, k-1.$$

The regression coefficient can be estimated using the matrices technique as:

$$\underline{Y} = \begin{bmatrix} 12.0 \\ 11.0 \\ 17.0 \\ 26.0 \\ 20.0 \\ 31.0 \\ 31.0 \\ 13.0 \\ 11.0 \\ 25.0 \\ 23.0 \\ 21.0 \\ 35.0 \\ 18.0 \\ 37.0 \\ 29.0 \\ 45.0 \\ 27.0 \\ 31.0 \\ 27.0 \\ 33.0 \\ 26.0 \\ 32.0 \\ 29.0 \end{bmatrix}, X = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \underline{\hat{\beta}} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix}, \qquad X'\underline{Y} = \begin{bmatrix} Y_{..} \\ Y_{1.} \\ Y_{2.} \end{bmatrix}$$

Where $x_0 = 1$ represent the coefficient of regressionβ_0. And other coefficient of the regression model $\underline{\beta}$can be estimated from:

$$(X'X)\underline{\hat{\beta}} = X'\underline{Y}$$

$$\begin{bmatrix} 24 & 8 & 8 \\ 8 & 8 & 0 \\ 8 & 0 & 8 \end{bmatrix} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} 610 \\ 161 \\ 199 \end{bmatrix}$$

The coefficient of regression can be estimated by solving this equation,

$$\underline{\hat{\beta}} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} \overline{Y}_{3.} \\ \overline{Y}_{1.} - \overline{Y}_{3.} \\ \overline{Y}_{2.} - \overline{Y}_{3.} \end{bmatrix} = \begin{bmatrix} 31.25 \\ -11.125 \\ -6.375 \end{bmatrix}$$

then the estimated regression equation takes the following form;

$$\hat{Y} = \beta_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2$$

$$\hat{Y} = 31.25 - 11.125X_1 - 6.375X_2$$

The hypothesis test is:

$H_0: t_1 = t_2 = 0$

If $t_1 - t_2 = 0$, then $t_3 = -t_1 - t_2 = 0$.

The sum of square of total is:

$$SST = \sum Y_{ij}^2 - \frac{Y_{..}^2}{nk} = 17270 - 15504.167 = 1765.833$$

But the sum square of regression coefficients model (β_0, β_1 and β_2) is:

$$R(\beta_0, \beta_1, \beta_2) = \beta' X' Y$$

$$= [31.25 \quad -11.125 \quad -6.375] \begin{bmatrix} 610 \\ 161 \\ 199 \end{bmatrix} = 16003.75$$

$$R(\beta_0, \beta_1, \beta_2) = 16002.45$$

$$R(\beta_0) = \frac{Y_{..}^2}{nk} = \frac{(610)^2}{24} = 15504.167$$

$$R(\beta_1, \beta_2 | \beta_0) = R(\beta_0, \beta_1, \beta_2) - R(\beta_0) = 499.583$$

$$SSE = SST - R(\beta_1, \beta_2 | \beta_0)$$

$$= 1765.833 - 499.583 = 1266.25$$

3.The multiple regression coefficients and analysis of variance output are shown in Table (4, and 5) using

SPSS program package.

Table .4 Regression Coefficients

Model	Unstandardized Coefficients		Standardized Coefficients	T	Sig.
	B	Std. Error	Beta		
(Constant)	31.25	2.746		11.378	0
x1	-11.125	3.884	-0.611	-2.864	0.009
x2	-6.375	3.884	-0.35	-1.641	0.116

a. Dependent Variable: Yi

Table . 5 Multiple Regression Analysis ANOVA[b]

Model	Sum of Squares	df	Mean Square	F	Sig.
Regression	498.583	2	249.292	4.131	.031[a]
Residual	1267.25	21	60.345		
Total	1765.833	23			

a. Predictors: (Constant), x2, x1

b. Dependent Variable: Yi

The analysis of variance for multiple regression analysis and completely randomized fixed model design gives the same outputs Tables (2) and (5). Analysis of variance (ANOVA) is similar to regression in that it is used to investigate and model the relationship between a response variable and one or more independent variables. However, analysis of variance differs from regression in two ways: the independent variables are qualitative (categorical), and no assumption is made about the nature of the relationship (that is, the model does not include coefficients for variables). In effect, analysis of variance extends the two-sample t-test for testing the equality of two population means to a more general null hypothesis of comparing the equality of more than two means, versus them not all being equal.

Conclusions:

In this articl we focused on delineating the difference between the cell means model and the effects model and reinforcing the notion that ANOVA fixed effects models can be analyzed using the general multiple regression model. A very simple explanation is that regression is the statistical model that is used to predict a continuous outcome on the basis of one or more continuous predictor variables. In contrast, ANOVA is the statistical model that is used to predict a continuous outcome on the basis of one or more categorical predictor variables. Regression and ANOVA have a lot in common:

First, both models are applicable only when you have a continuous outcome variable. A categorical outcome variable would rule out the use of either a regression model or an ANOVA model.

Second, the regression algorithm can use, which is based on the principle of least squares, to fit an ANOVA model. You don't have to use the least squares principle because there are other ways to produce the ANOVA model. But because least squares, the basis for regression models, also works for ANOVA models, some people consider the regression model to be the more general model. We can incorporate categorical predictors into a regression model by using indicator variables. An indicator variable is equal to one for a particular category and zero for the remaining categories. If you have a categorical predictor variable with k levels, then you can input (k-1) indicator variables (the last indicator is always redundant) in a regression program and effectively get the same results as an ANOVA model.

Third, the concept of partitioning variation into sums of squares (SS) in an ANOVA model also provides a nice way to examine complex regression models. In an ANOVA model, the total variation (total SS) is partitioned into variation between groups (between SS) and variation within groups (within SS). We can do the same sort of thing for a regression model, partitioning total variation into variation due to the model (model SS) and variation unexplained by the model (error SS).

Fourth, regression models and ANOVA models share many of the same diagnostic procedures (procedures used to examine the underlying assumptions). In particular, we can compute residuals in both models and the plots involving those residuals are often very helpful.

Fifth, in the case of multiple linear regression model independent variables may be correlated with each other and also with the dependent variable. But the analysis of variance beyond the problem of the correlated independent variables as long as deals with independent qualitative variables this makes calculations easier in analysis of variance comparing to regression analysis. Also, the connection between ANOVA and multiple regression produce can be utilliz from SPSS.

References:

Berger Martijn P. F., and KeeWengWong(2009), "An Introduction to Optimal Designs for Social and Biomedical Research", John Wiley and Sons.

DeCsterJamie(2006), "Testing Group Differences using T-tests, ANOVA, and Nonparametric Measures". University of Alabama.

Gary W. Oehlert (2010), "A First Course in Design and Analysis of Experiments", University of Minnesota.

Garavaglia Susan and Sharma Asha (1998), "A Smart Guide to Dummy Variables: Four Applications and a Macro". Dun & Bradstreet , Murray Hill, New Jersey 07974.

https://onlinecourses.science.psu.edu/stat502/node/217,"Alternative Regression Approach to Single-Factor ANOVA". Eberly College of Science, Pennsylvania State University. Last revised 8 / 7 / 2015.

Kleinbaum David, Kupper Lawrence, Nizam Azhar, Rosenberg Eli (2008), "Applied Regression Analysis and Other Multivariable Methods". Fourth edition THOMSON, Duxburg.

Mason Robert L., Gunst Richard F, Hess James L.(2003), "Statistical Design and analysis of Experiments With Applications to Engineering and Science. Second edition a John Willy and Sons.

Milliken George A., and Johnson Dallas E.(2009), "Analysis of Messy Data". Chapman &Hall.

Muhammad Tharwat Abdelmonem(2004), "Design and Analysis of Experiments". Egypt Angelo Library.

Neter John, Kutner Michael, Wasserman William, and Nachtsheim Christopher (1996), "Applied Linear Statistical Models, 4th edition, Illinois.

Rawlings John O., PantulaSastry G., and Dickey David A. (1998), "Applied Regression Analysis: A Research Tool", Second Edition. Springer- Verlag New York USA.

Skrivanek Smita (2009), "The Use of Dummy Variables in Regression Analysis". Principal Statistician, MoreSteam.com LLC.

Seltman Howard J. (2014), "Experimental Design and Analysis". College of Humanities and Social Sciences at Carnegie Mellon University.

Toutenburg, Helge Shalabh(2009), "Statistical Analysis of Designed Experiments". Third Edition. Springer New York Dordrecht Heidelberg London .

Vuchkov Ivan N. and Boyadjieva Lidia N.(2001), "Quality Improvement with Design of Experiments A Response Surface Approach".KluwerAcademic Publishers, London.