Global Journal of Science Frontier Research: F
Mathematics and Decision Sciences
Volume 18 Issue 4 Version 1.0 Year 2018
Type : Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals
Online ISSN: 2249-4626 & Print ISSN: 0975-5896

Regression Analysis of an Experiment with Treatment as a Qualitative Predictor

By Eze, Chinonso Michael, Asogwa, Oluchukwu Chukwuemeka & Onwuamaeze, Uchenna Charity
University of Nigeria

Abstract- This research work explores the indept use of multiple regression analysis in modelling the optimum weight gain of Broiler chicken as a function of initial weight and feed combination (treatment). The statistical relationship between the weight gain, initial weight and different feed combination was established using a regression model. The measurement on the initial weights was kept silent, allowing the use of analysis of variance technique to determine the error inherent in the study. Application of the model was done on the result of the experiment carried out on agritted breed of Broilers. The best feed combination (among the different feed combination considered) was determined by looking out for the combination that yields the optimum weight gain when the initial weight is known.

Keywords: *multiple regression; qualitative predictor.*

GJSFR-F Classification: *FOR Code: MSC 2010: 47N30*

REGRESSIONANALYSISOFANEXPERIMENTWITHTREATMENTASAQUALITATIVEPREDICTOR

Strictly as per the compliance and regulations of:

Notes

Regression Analysis of an Experiment with Treatment as a Qualitative Predictor

Eze, Chinonso Michael [α], Asogwa, Oluchukwu Chukwuemeka [σ] & Onwuamaeze, Uchenna Charity [ρ]

Global Journal of Science Frontier Research (F) Volume XVIII Issue IV Version I Year 2018

Abstract- This research work explores the indept use of multiple regression analysis in modelling the optimum weight gain of Broiler chicken as a function of initial weight and feed combination(treatment). The statistical relationship between the weight gain, initial weight and different feed combination was established using a regression model. The measurement on the initial weights was kept silent, allowing the use of analysis of variance technique to determine the error inherent in the study. Application of the model was done on the result of the experiment carried out on agritted breed of Broilers. The best feed combination (among the different feed combination considered) was determined by looking out for the combination that yields the optimum weight gain when the initial weight is known.

Keywords: multiple regression; qualitative predictor.

I. Introduction

In design and analysis of experiment, interest is centred on getting a design structure that will facilitate the collection of appropriate data for proper analysis. In an experimental design, some necessary considerations are made pertaining the scope of the study, the factors(controllable and uncontrollable) to be studied, suitable layout(design) and the experimental units(plots). These considerations are made prior to the conduct of the experiment and they avail the experimenter the opportunity to understand the aims of the experiment and possibly identify the appropriate technique for the actualization of the set aside aims.

Basically, the purpose of every experiment conducted is to ascertain the effects and the relationship between the dependent variable and some other variables usually known as the predictor variables (Udom, 2015). The ability of a particular analytical technique to actualize the purpose of ascertaining the effects of identified factors depends on the structure provided for the study. In other words, if the proper structure is not provided, valid effects may not be ascertained. When several treatments or treatment combinations are randomly applied to the experimental units, analysis of variance (ANOVA) technique emerges as one of the tools which can analyze the linear model(s) that represent the experiment situation. The ANOVA procedure attempts to analyze the variation in a set of responses and assign portions of this variation to each variable in a set of independent variables(Wackerly et, al, 2008). With the concept of pairwise comparison, the significant treatment effects among all the treatments under study can be determined.

Author α: (Corresponding author) Department of Statistics, University of Nigeria Nsukka. e-mail: chinonso.eze@unn.edu.ng

Author σ: Department of maths/comp.sc/stats/informatics, Federal University Ndufu-alike Ikwo, Ebonyi State, Nigeria.
e-mail: qackasoo@yahoo.com

Author ρ: Department of Statistics, University of Nigeria Nsukka. e-mail: uchenna.onwuamaeze@unn.edu.ng

On the other hand, the analysis of variance layout for the experiment can be re-arranged such that the observations can be analyzed using multiple linear regression model where the treatments or treatment combinations form part of the independent variables. Multiple linear regression is an extension of simple linear regression to allow for more than one independent variable (Mendenhall, 2003). So, we combine the treatment and the initial weight to form the two factors that affect the weight gain. Oftentimes, The treatments in experiments are not quantitative, thus, it becomes appropriate for them to be premised on an indicator variable like

Notes

$$z_i = \begin{cases} 1 \; if \; i^{th} \; treatment \; is \; applied \; to \; the \; j^{th} replicate \\ 0 \; if \; otherwise \end{cases} \quad (1)$$

so that they can form part of the information matrix. The model appropriate for such setting can be expressed in compact form as $y = X\beta + \varepsilon$, where $y = y_{ij}$ is the observed response of j^{th} replicate receiving i^{th} treatment, β is a vector that contains the treatment effects, X is the information matrix containing the indicator variables and ε is error column matrix. The task of predictive analysis is to obtain β such that the predicted value is close to the observed value as much as possible (Motoyama, 1978). However, the effects of the treatments cannot be estimated using $\beta = (X'X)^{-1}X'y$ since the first column (the coefficient of the constant in the model) in X is the sum of the second and subsequent columns, which are linearly independent. Also, $(X'X)^{-1}$ does not exist since X is not of full rank. However, if another one or more quantitative measurements are made on the experimental units and added to the model as an additional independent variable different from the indicator variables such that the first column in X is not the sum of the second and subsequent columns, $(X'X)^{-1}$ will exist and $(X'X)^{-1}X'y$ will be able to estimate the parameters β. Inferences on β can then be made to determine which effect is higher than the other.

In the area of weight gain analysis as it applies to other arears of life, meta-analysis has often been used. This involves statistical procedure that combines the results of multiple scientific studies. Nianogo, et, al (2017) adopted meta-analysis in studying the weight gain of prisoners during incarceration. A meta-analysis of eight studies combined showed an average weight gain of 0.43 (0.14, 0.72) lb/week. In all the studies they consulted, a high proportion (43% to 73%) of participants reported weight gain during incarceration. Thus, they recommended the incorporation of initiatives aimed at combating unhealthy weight developments in health promotion activities within prisons.

However, our interest is to first, analyze the result of the experiment in a regression format and make use of the ANOVA precedure to analyze the variation in the experiment and assign portions of this variation to each variable in a set of independent variables while relaxing the initial weight. Effort is made also to determine the feed combination that yields optimum weight gain so as to solve the problem of unimaginable increase in the cost of poultry feed stuff pointed out by Adene (2004) as the greatest source of dilemma in poultry industry. For other problems associated with Broiler feeding (see Berepubo et al (1995), Offiong and olumu (1980), and Akpodiete (2008)).

Global Journal of Science Frontier Research (F) Volume XVIII Issue IV Version I Year 2018

II. Design of the Experiment

The experiment was designed in a complete randomized design with one factor which is a combination of PKC based feed and bioactive yeast having four and three levels respectively.

The treatments in the experiment are represented as follows:

Notes

T_1 = 0.4g of bioactive yeast/kg X 15kg of PKC based feed
T_2 = 0.8g of bioactive yeast/kg X 15kg of PKC based feed
T_3 = 1.2g of bioactive yeast/kg X 15kg of PKC based feed
T_4 = 0.4g of bioactive yeast/kg X 20kg of PKC based feed
T_5 = 0.8g of bioactive yeast/kg X 20kg of PKC based feed
T_6 = 1.2g of bioactive yeast/kg X 20kg of PKC based feed
T_7 = 0.4g of bioactive yeast/kg X 25kg of PKC based feed
T_8 = 0.8g of bioactive yeast/kg X 25kg of PKC based feed
T_9 = 1.2g of bioactive yeast/kg X 25kg of PKC based feed
T_{10} = 0.4g of bioactive yeast/kg X 30kg of PKC based feed
T_{11} = 0.8g of bioactive yeast/kg X 30kg of PKC based feed
T_{12} = 1.2g of bioactive yeast/kg X 30kg of PKC based feed

Since interest is on the growth of broilers, measurements were taken on

1. The initial weight gain
2. The weight gain of the broilers

Table 1: Layout of the Experiment

	Replication						
TREATMENT	1	2	3	4	5	6	7
T1	X_{11}	X_{12}	X_{13}	X_{14}	X_{15}	X_{16}	X_{17}
T2	X_{21}	X_{22}	X_{23}	X_{24}	X_{25}	X_{26}	X_{27}
T3	X_{31}	X_{32}	X_{33}	X_{34}	X_{35}	X_{36}	X_{37}
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
T12	X_{121}	X_{122}	X_{123}	X_{124}	X_{125}	X_{126}	X_{127}

III. Methodology

Having stated earlier that one of the aims of this work is to develop an adequate model for predicting the weight gain of broilers given the initial weight and a specific feed combination, it is appropriate to analyze the data with multiple regression analysis technique.

Multiple regression analysis is a statistical evaluation of the relationship between one dependent variable and two or more independent variables. This technique provides an adequate mathematical model that explains the relationship between the variables under consideration. The general multiple regression model in a compact form can be written as

$$Y = \beta X + \varepsilon \tag{2}$$

Y is a column vector containing the weight gain of broilers considered in the experiment

β is a vector of the partial slopes or partial regression coefficients (the intercept or general constant inclusive) (see Mendenhall et, al, 2003)

X is n X k matrix of the independent (predictor) variables which in this case are, the feed combination and initial weight.

ε is the random error associated with the dependent variable Y

However, interest also is on examining the effects of different feed combination, thus, the analysis is done using an appropriate analysis of variance model. Here, we ignore the measurements on the initial weights and concentrate only on the weight gain and the feed combination effects. Since we are considering only one factor in this problem, which is the feed combination, we consider a one-way analysis of variance model given as

$$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij} \qquad (3)$$

Y_{ij} is the weight gain of jth broiler fed with ith feed combination

μ is the grand mean

α_i is the effect of the ith feed combination

ε_{ij} is the error associated with the weight gain of jth broiler fed with ith feed combination

With eqt (3) above, we can determine the significant effect of the different feed combination on the weight gain of broilers considered in the experiment.

Thirdly, a proper adjustment on the initial weight (covariate) can be made so as to ascertain the true effect of the different feed combination. This can be achieved through analysis of covariance

IV. Analysis and Results

As we know, statistical analysis- modeling and inference form the basis for objective generalizations from the observed data (Jammalamadaka & SenGupta, 2001). Here, we shall apply the methodologies given in the previous section on the weight gain data.

Consider rewritting the general multiple regression model in (2) above in a more explicit form.

$$Y_{ij} = \mu + \sum_{i=1}^{12} \alpha_i z_{ij} + \beta X_{ij} + \varepsilon_{ij} \qquad (4)$$

In contrast to quantitative predictor variables, quality predictor variables can be entered into a regression model through dummy or indicator variables (Mendenhall et, al, 2003).

Observing the fact that z_{ij} is a dummy (indicator) variable defined as eqt (1), eqt (4) transforms to

$$Y_{ij} = \mu + \alpha_i + \beta X_{ij} + \varepsilon_{ij}\,;\ i = 1,2,\ldots,12;\ j = 1,2,\ldots,7 \qquad (5)$$

The definition of the model components remains the same as in (2), with α_i being the effect on weight gain due to i^{th} feed combination.

Thus, the model can be expressed as $(\text{weight gain})_{ij}$ = general mean + $(\text{feed combination})_i$ + slope$(\text{initial weight})_{ij}$ + $(\text{error})_{ij}$.

Notes

Global Journal of Science Frontier Research (F) Volume XVIII Issue IV Version I Year 2018

Notes

a) Assumptions of the Model

In order to carry out a linear regression analysis on a set of data, it is reasonable to assume that the variables under consideration satisfy the following assumptions:

- The dependent variables as well as the error terms are normally distributed with mean zero and variance δ^2. That is $Y_{ij} \sim N(0, \delta^2)$.
- The variance of the dependent variables as well as the variance of the error terms are the same for all the populations under study (Homoscedasticity).
- The independent variables are fixed (they can be controlled by the experimenter).

These assumptions can satisfactorily be justified by the normality and constant variance test performed in the following section.

i. *Normality Test*

Maximum weight gain = 1.1
Minimum weight gain = -0.41

Range (R) = maximum – minimum = 1.1 – (-0.41) = 1.51

By the application of Sturge's rule, we obtain the number of classes (C) as

C = 1+3.322Log_{10}N; N is the total number of observations which in this case is 84.
Therefore, C = 1+3.322Log_{10}84 = 7.3925 (approximately 7)

The class size (S) is obtained by the ratio of the range to the approximate number of classes. That is $S = \frac{R}{C} = \frac{1.51}{7} = 0.22$

The mean and standard deviation of the observations are 0.335 and 0.2780 respectively.

Table 2: Frequency Table of the Weight Gain

C. interval	FREQ.	X	C. boundary	$z = \frac{x - \bar{x}}{s}$	$P_i = P(z)$	$Fe = NP_i$
-0.42 - -0.19	2	-0.305	-∞ – −0.185	≤ −2.30	0.0107	0.8988
-0.18 – 0.05	4	-0.065	-0.185 - -0.045	-2.30 - -1.37	0.0746	6.2664
0.06 – 0.29	39	0.175	-0.045 - -0.295	-1.37 - -0.14	0.359	30.156
0.3 – 0.53	24	0.415	-0.295 – 0.535	-0.14 – 0.72	0.3199	26.8716
0.54 – 0.77	7	0.655	0.535 – 0.775	0.72 – 1.58	0.1787	15.0108
0.78 – 1.01	6	0.895	0.775 – 1.015	1.58 – 2.45	0.05	4.2
1.02 – 1.25	2	1.135	1.015 - +∞	≥ 2.45	0.0071	0.5964
TOTAL	84					

With chi-square test statistic, the test was conducted and it yielded test value of 4.2568. At $\alpha = 0.01$ and (4-2-1=1) degree of freedom, the tabulated value is 6.63490. Based on the values above, it was concluded the weight gain observations are normally distributed.

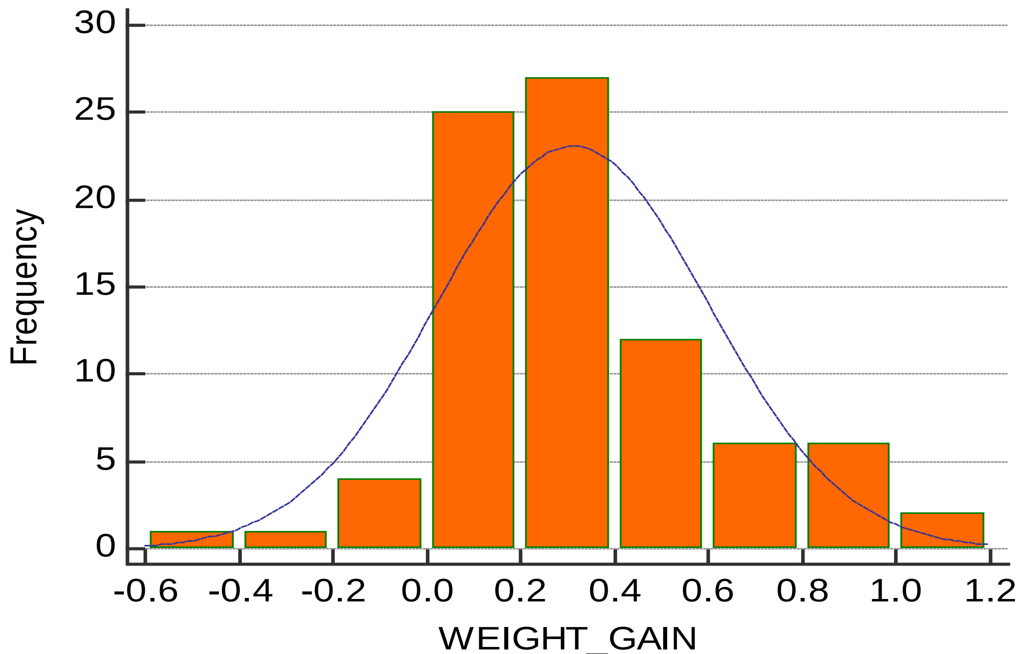

Figure 1: Normal Graph for the Weight Gain

ii. *Test For Constant Variance (Homoscedasticity)*

Another important assumption for the use of the model specified in section (3) is the assumption of constant variance. This assumption is important in that it makes the least square estimates of the model parameters to be linear unbiased estimates. In testing for constant variance, the most widely used procedure is the Bartlett's test. The procedure involves computing a statistic whose sampling distribution is closely approximated by the chi-square distribution with r-1 degree of freedom when random samples are from independent normal population.

Our interest is to show that the variances are the same across the groups. That is $\sigma_1^2 = \sigma_2^2 = \ldots = \sigma_{12}^2$. we made use of the chi-square distributed Test statistic below for the test

$$x^2 = 2.3026\frac{q}{c} \sim x_{r-1}^2$$

$$q = (N-r)Log_{10}S_p^2 - \sum_{i=1}^{r}(n_i - 1)Log_{10}S_i^2$$

$$c = 1 + \frac{1}{3(r-1)}[\sum_{i=1}^{r}(n_i - 1)^{-1} - (N-r)^{-1}]$$

$$S_p^2 = \frac{\sum_{i=1}^{r}(n_i - 1)S_i^2}{N-r}$$

Where N is the total number of observations, r is the number of populations, S_i^2 is the sample variance of the i[th] population, S_p^2 is the weighted average of the sample variance and n_i is the number of observations in i[th] population.

Notes

Global Journal of Science Frontier Research (F) Volume XVIII Issue IV Version I Year 2018

Table 3: Table of Sample Variances

$S_1^2 = 0.0092$	$S_5^2 = 0.1076$	$S_9^2 = 0.0635$
$S_2^2 = 0.0643$	$S_6^2 = 0.0369$	$S_{10}^2 = 0.0745$
$S_3^2 = 0.01123$	$S_7^2 = 0.0253$	$S_{11}^2 = 0.0369$
$S_4^2 = 0.0358$	$S_8^2 = 0.0679$	$S_{12}^2 = 0.1739$

Notes

$n_i = 7$

$$S_p^2 = (7-1)\frac{0.0092 + 0.0643 + \ldots + 0.1739}{84-12} = 0.0673$$

$$q = (84-12)Log_{10}(0.0673) - 6[Log_{10}(0.0092) + Log_{10}(0.0643) + \ldots + Log_{10}(0.1739)] = 7.463$$

$$C = 1 + \frac{1}{3(12-1)}\left[\frac{1}{6} + \frac{1}{6} + \ldots + \frac{1}{6} - \frac{1}{72}\right] = 1.0602$$

The test statistic value therefore, becomes

$$x_{cal}^2 = 2.3026\frac{7.463}{1.0602} = 16.2085$$

At 0.05 level of significance and r-1=11 degree of freedom, the tabulated value, $x_{11}^2; 0.05$ =19.675.

Since the Bartlett's test statistic value (16.2085) is less than the chi-square tabulated value (19.675), we conclude that all the twelve population variances are the same at 0.05 level of significant.

Let eqt (5) be presented in terms of general linear regression model as in eqt (2),

$$Y = XB + e$$

Where $\boldsymbol{Y}$ is an $n \times 1$ vector of the dependent variable (the weight gain), $\boldsymbol{X}$ is an $n \times m$ matrix of the independent variables which contains information about the initial weight and feed combination considered, $\boldsymbol{B}$ is an $m \times 1$ vector of the parameters (the general constant inclusive) and $\boldsymbol{e}$ is an $n \times 1$ vector of the model error. With ordinary least square approach, we estimate the parameters of the model such that the error function is minimized. To get an optimum set of parameters for the model, we consider the sum of square of the error component and differentiate with respect to the parameter set as follows:

Consider

$Y = XB + e$, the sum of square of the error term implies

$$\mathrm{SSE} = \mathrm{e}'\mathrm{e} = (\mathrm{Y} - \mathrm{X}\beta)'(\mathrm{Y} - \mathrm{X}\beta)$$

Opening the bracket, we have that

$$\mathrm{SSE} = \mathrm{Y}'\mathrm{Y} - \mathrm{Y}'\mathrm{X}\beta - \mathrm{X}'\beta'\mathrm{Y} + \mathrm{X}'\beta'\mathrm{X}\beta$$

But $\mathrm{Y}'\mathrm{X}\beta = \mathrm{X}'\beta'\mathrm{Y}$

Therefore,

$$\mathrm{SSE} = \mathrm{Y}'\mathrm{Y} - 2\mathrm{X}'\beta'\mathrm{Y} + \mathrm{X}'\beta'\mathrm{X}\beta$$

Differentiating partially the sum of square of the error with respect to the elements of the parameter vector implies

$$\frac{\partial SSE}{\partial \beta} = -2X'Y + 2X'X\hat{\beta} = 0$$

$$\rightarrow X'X\hat{\beta} = X'Y$$

Pre-multiply both sides by $(X'X)^{-1}$ to have $\hat{\beta} = (X'X)^{-1}X'Y$.

$X'X$ is an information matrix that contains the relationship (covariance) existing among the independent variables. Fixing the data considered in this research work into the formula derived above, the following estimated values of the model parameters are generated as presented in the table below

Notes

Table 4: Parameter Estimates

Parameter	Estimate	Parameter	Estimate
constant	0.0309	α_7	0.2420
α_1	0.1236	α_8	0.3685
α_2	0.0823	α_9	0.5871
α_3	0.3667	α_{10}	0.6592
α_4	0.2782	α_{11}	0.1157
α_5	0.3824	α_{12}	0.4183
α_6	0.3222	β	-0.0301

Therefore, the fitted model is

$Y_{ij} = 0.0309 + \alpha_i - 0.0301X_{ij}\,;\ \alpha_i's$ are presented in the table above.

b) *Goodness of Fit Test*

Having fitted a regression model to the set of the observations, it is appropriate to assess the robustness (goodness) of the fitted model. This test aims at knowing whether the true statistical relationship between the weight gain, the treatment effect (feed combination) and the covariate variable (initial weight) is reflected in the fitted model. Here, we carry out the test under the null hypothesis that each of the parameters $B = 0$ (the parameters are all equal to zero) against the alternative that at least one of the parameters is not equal to zero. Consider the analysis of variance table of the model given below.

Table 5: Analysis of Variance Table 1

Source of variation	DF	SS	MS	F	P
Feed comb. And initial weight	2	9.9420	4.97102	78.35	0.000
Error	81	5.139	0.06344		
Total	83	15.0811			

The p-value (0.000) of the test is very small (smaller than any significance level one can imagine). This means that the predictor variables add significant information to the prediction of weight gain. Therefore, we conclude that since none of the model parameters is statistically equivalent to zero, the true linear relationship among the variables is reflected and the model optimally mimic the observations. In addition to the test above, the coefficient of determination (R^2) was also considered to ascertain the proportion of the total variation in the weight gain that is explained by the initial weight and feed combination. The metric that captures the coefficient of determination is expressed as

$CD = \frac{SS_{feed\ comb\ .and\ initial\ weig\ ht}}{Total\ sum\ sum\ of\ square}$. From the Analysis of variance table 1,

$CD = \frac{9.9420}{15.0811} = 0.66$. This implies that 66% of the entire variation in the weight gain experiment is accounted for by the initial weight and feed combination, and the remaining 34% is attributed to the uncontrollable factors.

Notes

However, it is often necessary in a multiple regression model to analyse the seperate effect of the independent variables. This will give a proper ground for comparison between multiple regression model and classical analysis of variance model. In testing for the significance of fitting the feed combination effects after allowing for initial weight effect, we try to find the sum of square due to the "scanty" model

$$Y_{ij} = \mu + \beta X_{ij} + e_{ij} \quad (6)$$

Here, the relationship between the weight gain and only the covariate (initial weight) is examined.

The information matrix from the observations is given as

$$X'X = \begin{bmatrix} 84 & 138.55 \\ 138.55 & 402.47 \end{bmatrix} \text{ and } X'Y = \begin{bmatrix} 26.05 \\ 49.88 \end{bmatrix}$$

From the above, $(X'X)^{-1} = \begin{bmatrix} 0.0275 & -0.0095 \\ -0.0095 & 0.0057 \end{bmatrix}$

Thus, the estimated parameter $\hat{\beta} = (X'X)^{-1}X'Y = \begin{bmatrix} 0.2425 \\ 0.0368 \end{bmatrix}$.

The sum of square for regression (based only on the initial weight) from the "scanty" model above is expressed as $SS_{IW(only)} = \beta' Y' X = 8.1527$. Therefore, the sum of square for feed combination only (after allowing for feed combination) is the difference between the sum of square feed combination and initial weight generated earlier and the sum of square for regression (based only on the initial weight) generated from the scanty model. That is

$$SS_{FC(only)} = SS_{feed\ comb\ .and\ initial\ weig\ ht} - SS_{IW(only)}$$

Therefore, $SS_{FC(only)} = 9.9420 - 8.1527 = 1.7893$.

Now, we have succeded in breaking the total variation in the entire experiment into components that cause them. These information are presented in the comprehensive ANOVA table below which will be considered for further analysis.

Table 6: Analysis of Variance Table 2

Source of variation	DF	SS	MS	F-ratio	P-value
Initial weight	1	8.1527	8.1527	128.5915	0.000
Feed combination after allowing for initial weight	1	1.7894	1.7894	28.224	0.000
Error	81	5.139	0.0634		
Total	83	15.0811			

c) Test For Individual Effect of the Independent Variables

Presenting feed combination as a factor, we intend to examine the significance of the factor in the entire experiment. Let the feed combination be denoted as $\gamma_i; i = 1,2, \ldots, 12$. We test the null hypothesis $(H_0: \gamma_i = 0)$ against the alternative that atleast one of the levels of the feed combination is significant. The test statistic which is the ratio of the mean square feed combination after allowing for initial weight to the mean square error, follows F-distribution with $k-1$ numerator degree of freedom and $n-k-1$ denominator degree of freedom. The test supports the rejection of the null hypothesis at 0.05 level of significance if $F_{cal} > F_{tab}$.

From the analysis of variance table above, $F_{cal} = 28.22$. At 0.05 level of significance and specified degrees of freedom, the P-value of the test is 0.000. This suggests that the feed combinations have significant effects on the weight gain since the $P - valu(0.000) < 0.05$. Similarly, the initial weights have significant effect on the weight gain since it is obvious that the $P - value(0.000) < 0.05$. This gives the ground for subsequent section where interest is to analyze the true effects of the feed combinations after making adjustment for or relaxing the initial weight.

It is necessary to go on to break further, the coefficient of determination in section () into coefficient of determination for initial weight and coefficient of determination for feed combination. This gives expilicitly, the proportion of variation accounted for by each of the factors.

$$CD_{initial\ weig\,ht} = \frac{SS_{initial\ weig\,ht}}{SS_{Total}} 100\% = 54\%$$

Similarly,

$$CD_{feed\ combination} = \frac{SS_{feed\ combin\,ation}}{SS_{Total}} = 12\%$$

This implies that out of the 66% of the total variation explained by the two regressors, 12% is attributed to the feed and 54% is attributed to the initial weight. This results justify the belief that the weight gain of broiler chicken strongly, is a function of the initial weight.

d) Prediction

Here, we adopt the fitted model $Y = 0.0309 + \alpha_i - 0.0301X$ to predict the weight gain of broilers if their initial weights (X) are known and they are fed with a particular feed combination α_i. The essence is to ascertain which feed combination among all under consideration yields the maximum weight gain. Suppose that two initial weights $(X = 0.24\ and\ 1.32)$ are randomly sellected, with the estimated effects of the feed combination, predictions are made.

Table 7: Prediction

Feed combination	Effect (α_i)	Initial weight	Weight gain
T1	0.1236	0.24	0.147276
		1.32	0.114768
T2	0.0823	0.24	0.105976
		1.32	0.073468
T3	0.3667	0.24	0.390376

Notes

Notes

		1.32	0.357868
T4	0.2782	0.24	0.301876
		1.32	0.269368
T5	0.3824	0.24	0.406076
		1.32	0.373568
T6	0.3222	0.24	0.345876
		1.32	0.313368
T7	0.2420	0.24	0.265676
		1.32	0.233168
T8	0.3685	0.24	0.392176
		1.32	0.359668
T9	0.5871	0.24	0.610776
		1.32	0.578268
T10	0.6592	0.24	0.682876
		1.32	0.650368
T11	0.1157	0.24	0.139376
		1.32	0.106868
T12	0.4183	0.24	0.441976
		1.32	0.409468

In further analysis to determine the most significant among the feed combination, we keep the initial weight constant and concentrate on the feed combination. We tabulate the feed combination (yeast and PKC based feed) according to the level specified in the experiment, in a complete randomized design in table (1) and study the observations with the model

$$\begin{cases} Y_{ij} = \emptyset + \pi_i + \varepsilon_{ij}\,;\ \varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2) \\ \sum_{i=1}^{12} \pi_i = 0 \end{cases} \tag{7}$$

First, the estimated marginal mean of the weight gain due to the feed combination only is tabulated below

Table 8: Estimated Marginal Mean

Feed combination	Mean	Variance	Std. Error	95% Confidence interval
T1	-0.1006	0.1903	0.1649	-0.4295 to 0.2282
T2	-0.1228	0.1675	0.1547	-0.4314 to 0.1857
T3	0.1747	0.1533	0.1480	-0.1204 to 0.4698
T4	0.1363	0.1087	0.1246	-0.1121 to 0.3847
T5	0.2962	0.0773	0.1051	0.08672 to 0.5058
T6	0.3182	0.0717	0.1012	0.1164 to 0.5200
T7	0.2352	0.0661	0.09717	0.04149 to 0.4290
T8	0.3898	0.0715	0.1011	0.1882 to 0.5914
T9	0.6293	0.0806	0.1073	0.4153 to 0.8433
T10	0.7957	0.1454	0.1441	0.5084 to 1.0831
T11	0.3309	0.2432	0.1864	-0.04077 to 0.7026
T12	0.6384	0.2506	0.1892	0.2611 to 1.0156

With model (7), the total variation in the entire experiment was partitioned into the components that caused them. Eventhough the initial weight is a factor in the experiment, it was kept silent so as to examine the through effects of the feed combination.

Global Journal of Science Frontier Research (F) Volume XVIII Issue IV Version I Year 2018

Table 9: Analysis of Variance Table 3

Source of variation	Sum of Squares	DF	Mean Square	F-ratio	P-value
Feed combination	2.1532	11	0.1957	2.906	0.003
Error(other fluctuations)	4.8493	72	0.0674		
Total	7.0025	83			

The p-value (p=0.003<0.05) suggests that the effects of the different feed combinations are not the same. To determine which is different from the other, we embark on multiple comparison on the estimated marginal means due to the different feed combinations, using Turkey Honestly significant difference (Turkey HSD) test. This approach of multiple comparison is based on the studentized t-distribution. It has the ability of controlling type one error by taking into account the number of means that are being compared.

The test statistic for Turkey HSD test is expressed as

$$Q = \frac{|T_i - T_j|}{\left(MSE/m\right)^{1/2}} \sim t_{n-m};\alpha \qquad (8)$$

T_i is the estimated marginal mean of the i^{th} group and T_j is the estimated marginal mean of the j^{th} group. $m = 7$ is the number of observations in a group while $n = 84(7 \times 12)$ is the total number of observations across the group.

Here, we estimate the MSE as the mean of the estimated group marginal variances. That is $\frac{\sum_{i=1}^{12} s_i^2}{12} = 0.136$.

Using (8) above, the comparison test was conducted for different possible pairs and the P-value of each of the pairs are presented in table (10).

Table 10: The P-Values of the Comparison Test

	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12
T1	-	1.000	0.959	0.987	0.679	0.603	0.858	0.359	0.019	0.001	0.558	0.016
T2		-	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
T3			-	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
T4				-	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
T5					-	1.000	1.000	1.000	1.000	1.000	1.000	1.000
T6						-	1.000	1.000	1.000	1.000	1.000	1.000
T7							-	1.000	1.000	1.000	1.000	1.000
T8								-	1.000	1.000	1.000	1.000
T9									-	1.000	1.000	1.000
T10										-	1.000	1.000
T11											-	1.000
T12												-

V. Summary, Conclussion and Recommendation

In this work, we have been able to model the weight gain data with a linear multiple regression equation, taking the treatment in the experiment as a qualitative predictor. Also, we have been able to adopt the analysis of variance technique in assigning portions of variation in the entire experiment to each variable in a set of independent variables. The mean square error in the regression approach (0.0634) is smaller than that of one way analysis of variance approach (0.0674). This can be attributed to the fact that the presence of the initial weight which was carried along in the former analysis helped in reducing the totality of the error in the entire experiment. When the initial weight is ignored, information is lost. Consequently, the mean square

Notes

error shuts up with a difference of 0.004 (6.31% increment). This implies that the initial weight contributes 0.4% accuracy to the analysis.

The adjustment on the model succeded in making the true mean effects of T1 and T2 to be negative, with T10 having the highest effect. This result is in synch with the effects estimated earlier in terms of regression estimates where it was shown that T1 and T2 still have the most least effect on the subjects.

Notes

In table(10), it is revealed that significant difference only exist between the marginal means T1 and T9, T1and T10, and T1 and T12. In every other comparison, there is no significant difference. This does not mean that the effect of these non significant feed combinations are realy the same. Rather it means that there is no enough (convincing) evidence to justify that they are different. A cursory look at the pairs that are significantly different, reveals that the difference between T1 and T10 is the most significant. This goes on to show that T10 (0.4g of bioactive yeast/kg X 30kg of PKC based feed) has highest treatment effect since it has higher estimated marginal mean of the weight gain than every other treatment. This inference is in line with the postulation made earlier under the section for multiple linear regression where it was noted that T10 has higher effect since the ordinary least square estimate of it is higher than every other parameter within the effect space.

References Références Referencias

1. Adene, D. F. (2004). Poultry health and production principles and practices. *String-Horden publishers. (Nig.) Ltd.*
2. Akpodiete (2008). The effect of PKCbased feed and Bioactive yeast on the growth of Broilers. *An unpublished thesis, University of Nigeria, Nsukka.*
3. Berepubo, N. A. (1995). Inclusion rate and true metabolizable energy of palm kernel meal in broilers diet in humid tropical environment. *Journal of Applied Animal Resources.*
4. Jammalamadaka, S. R and SenGupta, A. (2001). Topics in circular statistics. *World scientific publishing co. pte. ltd.*
5. Kleseges, R.C, Kleseges, L. M, Eck, L. M, and Shelton, M. L. (1995). A longitudinal analysis of accelerated weight gain in preschool children. P*ubMed (Peadiatrics).;95(1): 126-30.*
6. Mendenhall, W., Beaver, R. J, and Beaver, B. M. (2003). Probability and statistics. *Brooks/cole, United State of America.*
7. Motoyama,T(1978): Comparison of three methods for prediction: the least square method, ridge regression and equal weighting. *IOWA University, Digital R*epository. Retrospective Theses and Dissertations.
8. Nianogo, R. A, Gebremariam, M. K, and Arah, O. A. (2017). Weight gain during incarceratio: systematic review and meta-analysis. P*ub Med (Obes Rev).; 19(1): 98-110.*
9. Offiong, S. A and Olomu, J. M. (1980). Economics of production of Broilers. *Nigerian Journal of Animal Production. 2: 68-73.*
10. Udom, A.U.(2015): Elements of Applied Mathematical Statistics. *University of Nigeria press limited.*
11. Wackerly, D. D., Mendenhall, W, and Scheaffer, R. L. (2008). Mathematical statistics with applications. *Thomson Learning, Inc.*

This page is intentionally left blank