THE USE OF SPSS AND GENSTAT COMPUTER SOFTWARE IN DATA ANALYSIS IN AGRICULTURE

U.C. NWAOGWUGWU[1] AND U. H. UDOH[2]

[1]Department of Animal Science, Abia State University, Umahia Campus, Nigeria;

[2]Department of Animal Science, Federal University of Uyo, Akwa Ibom State, Nigeria

[1]ugwumbaisaac@gmail.com

Date Submitted: 6 March, 2014; Date accepted: 28 August, 2014.

Abstract

The study covers steps involved in using SPSS and GenStat statistical software to analyse data from agricultural research especially in Animal Science. The objective is to highlight the use of the software in analysis of research data. Emphasis is laid on understanding designs of experiments which will serve as a guide to performing any analysis chosen. The experimental designs used were Completely Randomized Design, Randomized Complete Block Design and Latin Square Design. Some statistical procedures included in the study are ANOVA, paired sample t-test, Pearson Product Moment Correlation and regression analyses. In addition, repeatability estimation following ANOVA was vividly outlined. This study also slightly compared the effectiveness of the two software used. Although SPSS was used in most of the analyses, its weakness compared to GenStat in areas like standard error (SE) of mean and ANOVA for repeatability estimation was shown.

Keywords: Computer, statistics, data analysis, SPSS, GenStat

Introduction

The use of computer for statistical processing and analyses of research data is very important in Agriculture in general and in Animal Science in particular where empirical research works are carried out. Data obtained from an experiment give little or no information unless they are properly analysed and interpreted. However, as important as it is, many academics in this field lack the knowledge of data analysis using computer. Perhaps this may be either because they cannot operate a computer or do not know how to use the statistical application packages. This handicap often makes the researcher to resort to manual methods of analysis which, however, are more cumbersome, time consuming and less accurate. Consequently, the progress of our research works is greatly hampered while the quality of our academic outputs is compromised. Monotony and drudgery in the choice of research topics and projects often result. This study tends to highlight the use of statistical application packages in the computer analysis of empirical data using software like SPSS (SPSS, 2004) and GenStat (GenStat, 2008).

Materials and Methods

Requirements for computer analysis of data include proper knowledge of applied statistics, computer literacy and statistical application packages such as the Statistical Package for Social Sciences (SPSS), GenStat, Statistical Analysis Systems (SAS), Harvey's Least-Squares and Maximum Likelihood. The need to have understanding of experimental design and statistical procedures to be employed in an analysis cannot be overemphasized. Common designs include Completely Randomized Design (CRD), Randomized Complete Block Design (RCBD) and Latin Square Design (LSD). Analysis of variance (ANOVA), t-test, correlation, regression and Chi-square analyses are examples of common statistical procedures employed in data analyses. There are numerous others and the choice depends on the nature and design of the experiment being carried out. In some cases, data are transformed before they are analysed. Different types of data transformations include square root transformation for counts, arc-sine transformation for proportion, logarithmic transformation (Snedechor and Cochran, 1989; Obi, 2002).

Data for analysis of variance in completely randomized design (CRD)

Model for CRD relevant to the analysis: $Y_{ij} = \mu + G_i + \in_{ij}$ where Y_{ij} is j^{th} body weight observed on i^{th} genotype; μ, overall mean; G_i, fixed effect of i^{th} genotype and $\in_{ij}$ is random error, assumed to be independently, identically and normally distributed with zero mean and homogenous variance [iind $(0, \sigma^2)$].

Table1: Body weight (g) of three genotypes of chickens at 3 weeks of age

Replicate	Genotype		
	Naked neck	Arbor Acre	Crossbred
1	103.10	200.54	125.80
2	106.53	186.36	134.54
3	109.95	212.85	130.20
4	100.15	185.60	122.70
5	110.28	190.71	140.35

Data for analysis of variance in randomized complete block design (RCBD)

Model for RCBD relevant to the analysis: $Y_{ijk} = \mu + X_i + G_j + \in_{ijk}$ where Y_{ijk} is k^{th} body weight observed on j^{th} genotype of i^{th} sex; μ, overall mean; X_i, fixed effect of i^{th} sex, and $\in_{ijk}$, random error [iind $(0, \sigma^2)$].

Data for a two factor factorial experiment (3x3) in CRD

Model for CRD of two factor factorial experiment relevant to the analysis: $Y_{ijk} = \mu + E_i + P_j + (EP)_{ij} + \in_{ijk}$ where Y_{ijk} is k^{th} keel length of individuals fed ij^{th} combined energy and protein levels; μ, overall mean; $(EP)_{ij}$, interaction effect of energy and protein; $\in_{ijk}$, random error [iind $(0, \sigma^2)$].

Table 2: Body weight (g) of three genotypes of male and female chickens at 3 weeks of age

Sex	Genotype			Sex	Genotype		
	Naked neck	Arbor Acre	Crossbred		Naked neck	Arbor Acre	Crossbred
Male	103.10	200.54	125.80	Female	80.00	120.00	135.10
Male	106.53	186.36	134.54	Female	95.55	125.50	110.50
Male	109.95	212.85	130.20	Female	90.25	135.85	96.45
Male	100.15	185.60	122.70	Female	86.50	115.00	115.55
Male	110.28	190.70	140.35	Female	85.50	100.00	112.25

Data for Paired sample t-test analysis

T-test is computed by the formula: $[\bar{X}_1 - \bar{X}_2] \div \sqrt{(SP/n_1) + (SP/n_2)}$, where $\bar{X}_1$ and $\bar{X}_2$ are means of sample one and two respectively. SP is pooled variance, which is given as $[(n_1-1)S^2_1 + (n_2-1)S^2_1] + [n_1+n_2-2]$. S^2 is variance, given as $[\{\sum X^2_i - (\sum (X)^2/n)\} \div (n-1)]$ and n is number of observation.

Data for Pearson Product Moment Correlation analysis

The correlation coefficient, r used to compute the Pearsòn Product Moment Correlation is given by the formula: $r = [\{\sum XY - (\sum X \sum Y/n)\}] \div [\sqrt{[\sum X^2 - (\sum (X)^2/n)][\sum Y^2 - (\sum (Y)^2/n)]}]$

Table 3: Keel length of Lanvander guinea fowl fed 3 levels of energy (kcal/kg) and three levels of protein (%) at 16 weeks of age

Treatment	Rep	KL (mm)	EP	Rep	KL (mm)	EP	Rep	KL (mm)
E1P1	1	10.14	E2P1	3	10.50	E2P1	3	11.00
EIP1	2	10.29	E2P1	1	11.00	E2P1	1	10.00
E1P1	3	10.29	E2P1	2	11.40	E2P1	2	10.80
E1P2	1	9.83	E2P2	1	11.00	E2P2	1	10.17
E1P2	2	10.50	E2P2	2	10.00	E2P2	2	9.83
EIP2	3	10.17	E2P2	3	10.83	E2P2	3	77.00
EIP3	1	10.40	E3P1	1	8.93	E3P1	1	10.14
E1P3	2	10.00	E3P1	2	10.29	E3P1	2	9.86
E2P1	3	10.57	E3P1	3	10.14	E3P1	3	10.14

EP = Treatment combinations of energy and protein

Data for regression analysis (Linear, Exponential, Semi-log, Double log functions)

Regression Models: Simple Linear: $Y = b_o\ b_1X + e_i$; Exponential: $Y = b_0 + b_1LnX + ei$; Semi-log: $LnY = b_0 + b_1X + ei$; Double log: $LnY = b_0 + b_1LnX + ei$. In each model, Y is the dependent variable (egg number), X is the independent variable (age), b_0 is intercept, b_1 slope or regression coefficient and Ln is natural logarithm.

Number of	Location	
Markets visited	Aba	Umuahia
1	3.00	9.00
2	5.00	12.00
3	7.00	10.00

Table 5: Linear body parameters and biologic marker of Chinchilla rabbit at 12 weeks of age.

Rep	**Ear length (cm)**	**Neck length (cm)**	**White blood cell**
1	4.54	6.34	1500.00
2	5.00	7.50	1800.00
3	3.68	6.23	1400.00
4	5.67	7.90	1600.00

Table 6: Number of eggs laid by 20 hens at different ages (weeks).

Age (weeks) (X)	20	25	30	35	40	45	50	55	60
Egg number (Y)	60	100	110	120	130	140	130	135	125

Table 7: Four days collection of faecal sample (g) from 4 WAD male goats fed 4 different diets (A, B, C, D)

Goat number	Days of collection			
1	150.46	176.40	136.50	124.38
	(A)	(B)	(C)	(D)
2	164.65	168.95	140.35	125.64
	(B)	(C)	(D)	(A)
3	148.40	154.62	138.90	139.30
	(C)	(D)	(A)	(B)
4	135.34	164.00	149.00	127.50
	(D)	(A)	(B)	(C)

Data for analysis of variance in a 4x4 Latin Square Design (LSD)

Model for LSD $Y_{ijkl} = \mu + \beta_{1i} + \beta_{2j} + T_k + \in_{ijkl}$ where Y_{ijkl} is the observation made on each goat and day, fed each diet, μ is the overall mean, β_{1i} and β_{2j} are effects of individual goat (roll) and day of collection (column) respectively, $\in_{ijkl}$, random error [iind $(0, \sigma^2)$].

Data for estimation of repeatability

Repeatability, R is calculated by $(V_G + VE_P + VGE_P)/V_P$, where V_G is genetic variance, VE_P, permanent environmental variance and VGE_P, their interaction. The analysis of variance (ANOVA) model for repeatability is given as $Y_{ij} = \mu + T_i + B_j + e_{ij}$, where Y_{ij} is the record of the j^{th} bird in the i^{th} time of measurement, μ, population mean, T_i, fixed effect of i^{th} time of measurement, B_j, random effect of j^{th} bird and e_{ij}, random error. Following ANOVA which yields variance components, repeatability can be estimated with the formula: $R = \sigma^2_b/[\sigma^2_b + \sigma^2_e]$, where σ^2_b is variation between birds and σ^2_e, error variance, i.e. variation between time of measurement (month) within birds. σ^2_b estimates all the variance due to permanent parts of the records. σ^2_e is the error mean square (MSe) and σ^2_b is $(MS_b - MSe)/k1$, where MS_b is the mean square of birds and k1 is number of months.

Table 8: Monthly egg production of 20 pullets over a period of 5 months

Bird No.	Month					Bird No.	Month				
	1	2	3	4	5		1	2	3	4	5
1	10	5	15	8	8	11	20	20	22	19	11
2	15	17	18	16	10	12	26	22	19	15	14
3	5	10	16	7	18	13	22	26	12	21	20
4	19	17	12	9	112	14	8	11	17	11	13
5	18	9	17	16	18	15	19	22	19	14	17
6	15	21	14	15	21	16	19	20	22	21	20
7	10	20	12	15	22	17	18	4	16	18	15
8	8	9	15	11	12	18	25	23	22	18	22
9	6	15	11	9	14	19	18	19	13	18	16
10	12	15	22	8	12	20	12	16	18	18	4

Source: Ibe (1998)

Table 9: Inputting data into SPSS and GenStat spreadsheets for ANOVA in CRD

Genotype	BWT (g)	Genotype	BWT (g)
1	103.10	3	125.80
1	106.53	3	134.54
1	109.95	3	130.20
1	100.15	3	122.70
1	110.28	3	140.35
2	200.54		
2	186.36		
2	212.85		
2	185.60		
2	190.71		

Methods of Data Analysis

Analysis of variance in CRD: Table 9 shows how to input data from CRD into SPSS or GenStat spreadsheet for ANOVA. The three treatments – Naked neck, Abboica and Crossbred are coded 1, 2 and 3, respectively. However, with the GenStat, the data may be input in Excel spreadsheet and consequently imported into in GenStat. Locate SPSS application on the programme list and double click to open it. Input the data in table 7 into the SPSS Data Editor. Click Variable View on the SPSS task bar and label Genotype and Body weight (g) under the Name column. 1, 2 and 3 can be defined by clicking the row under Values.

Click the Data View to go back to SPSS Data Editor. On the menu bar, click analyze, drag the mouse to compare means, follow the right arrow and move down and click One-Way ANOVA.

Click the Data View to go back to SPSS Data Editor. On the menu bar, click analyze, drag the mouse to compare means, follow the right arrow and move down and click One-Way ANOVA. Transfer Genotype to the Factor and body weight (g) to the Dependent List. Click Post Hoc, Duncan and Continue. Click Options, Descriptive and Continue and Ok.

Analysis of variance in RCBD. Create extra column for block and follow these steps for analysis by clicking or dragging: Analyze, General Linear Model, Unvaried; transfer block and genotype to the Fixed Factor (s) and body weight to the Dependent Variable; click Model, Custom and under Build Term (s), Main effects; transfer genotype from Factors & Covariate to Model; click Post Hoc, transfer Genotype from Factor (s) to Post Hoc Tests, click Duncan, Continue; Click Option, transfer genotype and body weight from Factor (s) and Interaction to Display Means for, Click Descriptive statistics, Continue, Ok.

Table 10: Inputting data into SPSS for ANOVA in RCBD.

Block (Sex)	Genotype	BWT (g)	Block (Sex)	Genotype	BWT (g)
1	1	103.10	2	1	80.00
1	1	106.53	2	1	95.55
1	1	109.95	2	1	90.25
1	1	100.15	2	1	86.50
1	1	110.28	2	1	85.50
1	2	200.54	2	2	120.00
1	2	186.36	2	2	125.50
1	2	212.85	2	2	135.85
1	2	185.60	2	2	115.00
1	2	190.70	2	2	100.00
1	3	125.80	2	3	135.10
1	3	134.54	2	3	110.50
1	3	130.20	2	3	96.45
1	3	122.70	2	3	115.55
1	3	140.35	2	3	112.25

Importing data from Excel into GenStat for ANOVA of 3x3 factorial experiments in CRD.

It is important to note that in the GenStat package, all treatments or independent variables are called factors and MUST be converted to factors, whereas all parameters, response or dependent variables are called variates. The data in table 11 are input into Excel spreadsheet. GenStat is opened, and using the Excel Import Wizard on the GenStat formatting tool bar, the table is transferred into GenStat active spreadsheet. If the two treatments were not converted to factors during the importation process, this can be done by highlighting and right-clicking Energy and Protein on the Active

Spreadsheet, and clicking convert to factors. A red indicator like a big cylindrical exclamation mark (!) shows on the row besides the two factors. The analysis is proceeded by performing (clicking) the following steps: start (on the menu bar), analysis of variance (drag down General Analysis of Variance), Two-Way ANOVA (no blocking). The Analysis of variance windows then opens. Under the Available Data dialogue box, double click to transfer KL (mm) to the Y- variate dialogue box. Transfer the factors by double clicking from under the Available Data dialogue box to the Treatment1 and Treatment2 dialogue boxes respectively by clicking on each Treatment dialogue box first. Click on options to open ANOVA Options windows and tick %cv and LSD. Click OK and OK to complete the analysis. The result is then accessed on the Output windows.

Table 11: Importation of data from Excel into GenStat for ANOVA of 3x3 Factorial Experiment in CRD

	Energy	Protein	KL (mm)	Energy	Protein	KL (mm)
E1		P1	10.14	E2	P2	10.83
E1		P1	10.29	E2	P3	8.93
E1		P1	10.29	E2	P3	10.29
E1		P2	9.83	E2	P3	10.14
E1		P2	10.50	E3	P1	11.00
E1		P2	10.17	E3	P1	10.00
E1		P3	10.40	E3	P1	11.00
E1		P3	10.00	E3	P1	10.00
E1		P3	10.57	E3	P1	10.80
E2		P1	10.50	E3	P2	10.17
E2		P1	11.00	E3	P2	9.83
E2		P1	11.40	E3	P2	77.00
E2		P2	11.00	E3	P2	10.14
E2		P2	10.00	E3	P2	9.86

T-test analysis: The data in Table 4 are input in that fashion into the SPSS for T-test analysis and the two locations defined. Analyze, Compare Means, Paired Samples T Test are clicked or traced. Highlight and transfer Aba and Umuahia into the Paired Variables and click OK to performer the analysis.

Correlation Analysis: Data in Table 5 are input into the SPSS and the variable, ear length, neck length and white blood cell are defined properly. Click analyze, drag to correlate and click Bivariate. Highlight El, NL and WBC and transfer them to the variable box and click OK run correlation. The result is shown in the output. For good result of correlation analysis, number of replication must not be less than three.

Regression Analysis: Regression analysis of age or independent variable (X) is performed on Number of egg or dependent (Y) using four regressions functions – Linear, exponential, semi-log and double log. Apart from the linear function which needs not any transformation, data for the other functions must be transformed to Lin (Ln).

The linear function is run in SPSS. The variables, X and Y are defined by clicking the variable view. To complete the analysis, proceed to click data view and analyze, drag the mouse to regression and to linear and click. The analyses of the other functions

must follow data transformation which is performed on GenStat or SPSS package. With the data in the SPSS spreadsheet, perform the following steps for Ln transformation: click transform on the menu bar, click compute to get compute variable windows, under function group, click Arithmetic, under Functions and Special Variables, double-click Ln, to transfer it to the Numeric Expression dialogue box. Exponential function requires that Ln transformation of Y variable should be performed. In Semi-log function, only X variables is transformed while in double log function, both X and Y variables are transformed. Double-click whichever one to transfer it to the Numeric Expression dialogue box. Then type the variable for transformation in the Target Variable dialogue box and click OK. The original untransformed data should be resaved in another column in the SPSS Data Editor to avoid it being replaced by the transformed variable.

Inputting data into SPSS for ANOVA in LSD

Performing ANOVA in Latin Square Design demands that the data should be arranged in a cyclic or rotational fashion (Table 11). Rows, R (animal number), Columns, C (days of collection) and Treatment, T (A, B, C, D) must be created before such analysis is performed using the data in Table 10. The numbers: 1, 2, 3, and 4 code for the actual levels of C and T. The factors, R, C and T are defined, labelled where necessary and coded in SPSS Variable Viewer Data Editor. Follow the steps outlined for the ANOVA in RCBD. However, the means with their standard errors are analysed separately using the Analyze, Compare Means and Means. Options are clicked so that Std. Error of Means can be selected.

Table 12: Method of inputting data into SPSS in LSD for ANOVA

R	1	1	1	1	2	2	2	2	3	3	3	3	4	4	4	4
C	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
T	1	2	3	4	2	3	4	1	3	4	1	2	4	1	2	3
QF (g)	150.46	176.4	136.50	124.38	164.65	168.95	140.35	125.64	148.40	154.62	138.9	139.3	135.34	164	149.	127.5

Data in Table 13 should be a continuous column when input into the software. Input the data into the SPSS Data Editor. On the variable view page, type and define EN, M and NE as Bird number, Month, and Number of egg. Click analyze on the menu bar and drag the mouse to General Linear Model and then select and click Univariate to choose the type of analysis transfer by NE to Dependent Variable, M to Fixed Factor and BN to Random Factor. Define the Model by clicking Model and Custom and transfer the factors into the Custom box. Click Continue and OK. In GenStat spreadsheet, input the data for repeatability and convert BN and M to factors. Click stats, drag down to Mixed Models (REML) and locate and Linear Mixed Models. Transfer NE to Y-Variate, M to Fixed Model and BN to Random Model. Click Options to open Linear Mixed Model Option and tick the relevant options. Click OK.

Results and Discussion

Tables 14 and 15 show the results of ANOVA from SPSS (2006) and GenStat (2008) using data in Table 3. SPSS result shows the standard errors of individual treatment means (SE), Standard error of difference of means (Sd) incidentally called the standard error (SX or SEM) and means separation using Duncan Multiple Range Test. GenStat result shows the standard error of mean (SX), F-LSD and Standard error of difference (S_d) of means, which are all derived from the ANOVA table (Table 2.2.1). The formulas for computing SE = $S^2/\sqrt{n}$; $S_X = \sqrt{Se^2/r}$; $S_d = \sqrt{2S^2e/r}$; F-LSD = $t_{(0.05)}$, edf x $\sqrt{S^2e/r}$, where S = variance, n = number of observation, r = number of replicate, edf = error degree of freedom. The result shows that Arbor Acre had the greatest significant ($P<0.05$) body weight (195.21± 5.15).

Table 13: Method of inputting egg production data in SPSS or GenStat for ANOVA and consequent repeatability estimation

BN	M	NE	BN	M	NE	BN	M	NE	BN	M	NE
1	1	10	6	1	15	11	1	20	16	1	19
1	2	5	6	2	21	11	2	20	16	2	20
1	3	15	6	3	17	11	3	22	16	3	22
1	4	8	6	4	15	11	4	19	16	4	21
1	5	8	6	5	21	11	5	11	16	5	20
2	1	15	7	1	10	12	1	26	17	1	18
2	2	17	7	2	20	12	2	22	17	2	4
2	3	18	7	3	14	12	3	19	17	3	16
2	4	16	7	4	15	12	4	15	17	4	18
2	5	10	7	5	22	12	5	14	17	5	15
3	1	5	8	1	8	13	1	22	18	1	25
3	2	10	8	2	9	13	2	26	18	2	23
3	3	15	8	3	12	13	3	12	18	3	22
3	4	7	8	4	11	13	4	21	18	4	18
3	5	18	8	5	12	13	5	20	18	5	22
4	1	19	9	1	6	14	1	8	19	1	18
4	2	17	9	2	15	14	2	11	19	2	19
4	3	16	9	3	15	14	3	17	19	3	13
4	4	9	9	4	9	14	4	11	19	4	18
4	5	12	9	5	14	14	5	13	19	5	16
5	1	18	10	1	12	15	1	19	20	1	12
5	2	9	10	2	15	15	2	22	20	2	16
5	3	12	10	3	11	15	3	19	20	3	18
5	4	16	10	4	8	15	4	14	20	4	18
5	5	18	10	5	12	15	5	17	20	5	4

BN=Bird no. M=Month, NE= Number of egg

Table 14: Mean body weight of three genotypes of chickens (SPSS Result from CRD)

Parameter	Genotype		
	Naked neck	Arbor Acre	Crossbred
BWT (g)	106.00 ± 1.96^{C}	195.21 ± 5.15^{a}	130.72 ± 3.13^{b}

^{a,b,c} Means in the same row having different superscripts different significantly ($P<0.05$)

Table 16 presents the multiple comparisons of the means of the three genotypes with the common standard error (7.89). This value is actually the standard error of difference and not the LSD, and it is derived from the AVONA table. This is ascertained when GenStat is used to analyse the same data (Table 17). This is one of the limitations of SPSS.

Table 16: Multiple comparisons of treatment means showing standard error of difference mistakenly called standard error of mean from SPSS Result

Multiple Comparisons

Dependent Variable: BWT

	(I) Genotype	(J) Genotype		Std. Error	Sig.
LSD	1.00	2.00	-60.4590*	7.88818	.000
		3.00	-25.5630*	7.88818	.003
	2.00	1.00	60.4590*	7.88818	.000
		3.00	34.8960*	7.88818	.000
	3.00	1.00	25.5630*	7.88818	.003
		2.00	-34.8960*	7.88818	.000

Based on observed means.

*. The mean difference is significant at the .05 level.

Table 17: Mean body weight of three genotypes of chickens (GenStat Result from RCBD)

Parameter	Genotype			$LSD_{0.05}$	$S_{\bar{d}}$
	Naked neck	Arbor Acre	Crossbred		
BWT (g)	96.78	157.24	122.34	16.21	7.89

Table18: Effect of energy and protein on keel length (mm) of guinea fowls

PROTEIN (%)	ENERGY (kcal/kg)			**Mean**
	1	2	3	
1	10.24	10.17	16.32	**17.80**
2	10.97	10.61	9.79	**11.54**
3	10.60	9.23	10.05	**10.71**
Mean	**18.59**	**10.90**	**10.55**	**13.35**

LSD For Energy (E) = 0.60^{NS} LSD for Protein (P) = 0.60* LSD for interaction of E.P = $1.05^{NS\ NS}$ =Not significant, *= Significant

The result of Table 18 shows that only the effect of protein on Keel length (mm) was significant ($P<0.05$); energy and interaction of energy and protein did not give any significant effect on the keel length of the guinea fowls. Protein 1 which influenced highest mean of keel length (17.80 mm) should be recommended for this important growth parameter.

The result of the paired samples T-test presented in Table 13 indicates that there is a significant difference ($P<0.05$) between the number of foetuses discovered from cows slaughtered at Aba and Umuahia. More of such activities take place in Umuahia than in Aba.

Table 19: Comparing the number of foetuses discovered from cows slaughtered at Aba and Umuahia

Parameter	Location	
	Aba	Umuahia
No. of foetus	5.00±1.15	10.33 ± 1.53a

[a,b] Means in the same row having different superscripts differed significantly (P<0.05)

Table 15 shows that there is a non-significant (P>0.05) positive increase in white blood cell with a corresponding increase of the body linear parameters. Therefore, the physical linear growth may give information about the biologic marker which aids in selection for certain traits such as resistance to certain diseases.

Table 20: Correlation between WBC and Linear Body Parameters (EL, NL) of Chinchilla rabbit at week 12

	EL (cm)	NL (cm)
WBC	0.66 [NS]	0.75 [NS]

[NS] = Not significant (P>0.05)

Result of regression analysis of egg number on age of hens.

In all the four functions tested, regression was significant (P<0.05) and the R2 was high above 50%, indicating that eggs laid actually depended on age in weeks, although this statement is true to some extent. For instance, in linear function, when age increases by one unit egg laid increases by 1.42 unit at a significant rate (P<0.05) and the coefficient of variation (R^2) of 0.62 shows that age in weeks contributed 62% among the total factors that influenced egg laying. It is better to use functions other than linear to run regression analysis. For instance, 60 in linear regression imply that at 0 weeks the number of eggs that will be laid is 60, but this is impossible because at that age the chicks have not even hatched.

Table 21: Regression of egg number on age of hens

Function	Regression equation	R^2	Standard error of the estimate	Sig
Linear	EN = 60.00 + 1.42Age	0.62	16.21	0.012*
Exponential	Log EN = 0.07 + 0.003 Age	0.96	0.01	0.000**
Semi-log	EN= 92.76 + 132.80 Log Age	0.76	13.02	0.002**
Double log	Log EN = -0.25 + 0.28 Log Age	0.998	0.002	0.000**

* = Significant at 5% ; ** = Significant at 1%

Table 22 shows the results the effect of diet on the quantity of faeces collected from 4 WAD goats. In Latin Square Design, the effects of rows and columns are not important to the researcher. They are therefore not included in result.

Table 22: Result of ANOVA in Latin Square Design

Quantity of faeces (g)	DIET			
	A	B	C	D
	138.67± 8.18	157.34 ± 8.22[a]	151.28 ±9.48[ab]	144.75 ± 6.28[bc]

[a,b,c,d] Means in the same row having different superscripts are significantly different (P<0.05).

Table 23: SPSS ANOVA table for estimating repeatability of egg production of 20 pullets over a period of 5 months

Source		Type III Sum of Squares	Df	Mean Square	F	Sig.
Intercept	Hypothesis	23623.690	1	23623.690	346.626	.000
	Error	1294.910	19	68.153(a)		
M	Hypothesis	49.360	4	12.340	.774	.545
	Error	1211.040	76	15.935(b)		
BN	Hypothesis	1294.910	19	68.153	4.277	.000
	Error	1211.040	76	15.935(b)		

a MS(BN) BN=bird number, M= Month

b MS(Error)

ANOVA for estimation of repeatability of egg production

In the analysis of variance using data in Table 12, SPSS includes only the error mean square, which in this analysis, is 15.93. Hence to estimate repeatability, R, the variance component of dams (variation between individuals birds) i.e. BN, will first be calculated using $\sigma^2_b = (MS_b - MSe)/k1$, where MS_b = mean square of dams, MSe= Error Mean Square, k1= Number of measurements (M) (Ibe, 1998). That is $\sigma^2_b = (68.153-15.935)/5 = 10.44$. Consequently, $R = \sigma^2_b /(\sigma^2_b + \sigma^2_e) = 10.44/(10.44+15.93) = 0.39$. However with GenStat, the value of σ^2_b is included, so that R can be estimated without having to calculate σ^2_b first.

GenStat result of variance component of dam and error mean square

*** Estimated Variance Components ***

Random term	Component	S.e.
Bird no.	10.44	4.45

*** Residual variance model ***

Term	Factor	Model (order)	Parameter	Estimate	S.e.
Residual	Identity		Sigma2	15.93	2.58

Conclusion

Proper analysis of research data requires good understanding and application of experimental designs, models and formulae for computation of statistics. Speed and accuracy of these analyses can be achieved with the use of computer aided software. This paper has tried to show some practical steps involved in performing common analyses in agriculture, particularly in animal science. The list of software used here is not exhaustive, but the understanding of their usage is basic and can be applied to other software.

References

GenStat (2008). GenStat discovery edition 3. GenStat release 7.2 DE (PC/Windows XP).

Ibe, S.N. (1989). An Introduction to Genetics and Animal Breeding (1st Ed). Ikeja, Longman.

Obi, I.U. (2002). Statistical Methods of Detecting differences Between Treatment means and Research Methodology Issues on Laboratory and Field Experiments.

Snedechor, G.W. and Cochran, W.G. (1989). Statistical Methods (8th Ed).Ames, Iowa, U.S.A.: Iowa State University Press.

SPSS (2004). Statistical package for Social Sciences. SPSS Inc. Version 13.0 for Windows.