Journal of Natural and Medical Sciences (JNMS)
vol. **15** (2) 2014
ISSN 1858-6805
e-ISSN 1858-6813

SUST
Journal of Natural and Medical Sciences

عمـــادة البحـــث العلمـــي
DEANSHIP OF SCIENTIFIC RESEARCH

Journal homepage: http://journals.sustech.edu/

Use of Principal Components Analysis to Study and Analyze the Animal Resources in North Kordofan State

Mohammed Osman Mekki[1*] and Adil Musa Younis[2]
1*Department of computational Statistics – College of Science & Statistics ,Kordofan University, Sudan. Email:mogtaba3_3@yahoo.com
2- Department of Applied Statistics – College of Science, Sudan University of Science and Technology, Sudan.
RECEIVED: 3/8/2013 ACCEPTED: 12/5/2014

ABSTRACT
This study examined factors that affecting the livestock population density in North Kordofan State. A questionnaire was designed gathering information's related to that topics from 758 livestock owners in 2012. Preliminary descriptive statistics were used at early stage explained the general features and percentage of livestock variables. Followed by further analysis involved factor analysis and principal components, to draw the important factors affects livestock population. The results showed that twelve factors were extracted, which explain 63.12% of the total variance. The study concluded that the main factors affecting livestock animal health, in general particularly, were water resources and vaccinations campaigns.

المستخلص
في هذه الورقة تمت دراسة العوامل التي تؤثر علي الثروة الحيوانية في ولاية شمال كردفان ، تم أخذ عينة حجمها 758 من أصحاب الماشية وتم أستخدام الأحصاء الوصفي ومن ثم تم أستخدام التحليل العاملي والمكونات الأساسية ، تم التوصل إلي أثني عشر عامل تفسر 63.12% من جملة التباين.
خلصت الدراسة إلى أن أهم العوامل التي تؤثر علي الثروة الحيوانية هي صحة الحيوان ومصادر المياه وحملات التطعيم.

KEYWORDS: principal components, Factor analysis.

INTRODUCTION

The Objective of the study is to examine the underling structure of the animal resources in North Kordofan State. Animal resources and agriculture were being the backbone of the country economy before oil explores, therefore it can rise again as a strong alternatives if we know the good position of the country in this field beside the adequate factors that can easily lead to success in this sector, especially if we recognize the raising demand on animal products. In previous studies, following the same statistical method, some of the results were as follow: Rabie Ragab Sadek (1984) Studies on genetic and environmental factor affecting the milk production of the Egyptian buffalo He noticed that seven factors affecting of the milk production: season of calving, yearof calving, farm, location period, age of first calving, service period and calving interval explained 53.8% of the total variance.(1) Ahmed Hassan Mohamed Haridy (2010) studied some factors affecting the acquired immunity in chickens. He noticed that seven factors affecting of the milk production explain 61.7% of the total variance. (2)

Principal components analysis and factor analysis

principal components analysis and factor analysis are methodologies with the goal of reducing a large number of variables into a limited number of dimensions referred to as components or factors which are able to account for all or most of the variability in the original [3-5]. The (PC) analysis was used to transform the given set of variables, X_1, X_2 …X_k, into a new set of composite variables that were orthogonal to or uncorrelated with each other. The objective was to identify groups of inter-correlated variables. [6,7] Principal Component Analysis (PCA) is the general name for a technique which uses sophisticated underlying mathematical principles to transforms a number of possibly correlated variables into a smaller number of variables called principal components. The origins of (PCA) lie in multivariate data analysis; however, it has a wide range of other applications. PCA has been called, one of the most important results from applied linear algebra [8] Let $X \in R^p$ be a random vector with mean μ and covariance matrix $\sum$ (note that we don't make any assumptions about the distribution of X). Then the principal component transformation is the transformation:

$X \rightarrow Y = \Gamma^{/}(X - \mu)$ where Γ is the orthogonal matrix consisting of the standardized eigenvectors corresponding to the eigenvalues $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \ldots\ldots \geq \lambda_p$ of $\sum$. Thus, $\sum = \Gamma\Lambda\Gamma'$, or equivalently, $\Gamma'\sum\Gamma=\Lambda$.

Given $X = (X_1, \ldots, X_p)'$

We call $a^{/}Xa$ standard linear combination (SLC) if $\sum a_i^2 = 1$

Find the SLC $a^{/}_{(1)} = (a_{11}. \ldots, a_{p1})$ so that $Y_1 = a^{/}_{(1)} X$ has maximal variance

Find the standard linear combination SLC $a^{/}_{(2)} = (a_{12}, \ldots, a_{p2})$ so that $Y_2 = a^{/}_{(2)}X$ has maximal variance, subject to the constraint that Y_2 is uncorrelated to Y_1.

Find the standard linear combination SLC $a^{/}_{(3)} = (a_{13}, \ldots, a_{p3})$ so that $Y_3 = a^{/}_{(3)}X$ has maximal variance, subject to the constraint that Y_3 is uncorrelated to Y_1 and Y_2 , etc...

Properties of PCA

$E(Y_i) = 0$

$Var(Y_i) = \lambda_i$

$Cov(Y_i, Y_j) = 0$ if $i \neq j$

$Var(Y_1) \geq Var(Y_2) \geq \ldots \geq Var(Y_p)$

No standard linear combination (SLC) of X has a larger variance than λ_1 the variance of the first principal component.

Limitations of PCA

The directions with largest variance are assumed to be of most interest. We only consider orthogonal transformations (rotations) of the original variables. (Kernel PCA is an extension of PCA that allows non-linear mappings). PCA is based only on the mean vector and the covariance matrix of the data. Some distributions (e.g. multivariate normal) are completely characterized by this, but others are not. Dimension reduction can only be achieved if the original variables were correlated. If the original variables were uncorrelated, PCA does nothing, except for ordering them according to their variance. PCA is not scale invariant.

Materials and Methods

This study was conducted in north Kordofan State at 2012. Questionnaires were used to collect data from owner of livestock to study the factors affecting of livestock population. The Questionnaires contained: owner gender, owner age, the owner family member, the breeding type, the breeding purpose, education level of the owner, the owner main professional, the owner secondary professional, the owner professional experience, breeding farm area of the owner, animals vaccinated or not, the number of vaccinated animal per year of the owner, the number of male slaughtering of animal per year, the number of female slaughtering of animal per year, the died of animal per year, the export animal per year, the animal commercial domestic per year, the animal grazing hours/day, the number of animal labor personnel, the others reasons for animal decrement, the numbers of animal vaccinated against diseases per year, the sources of drink water of the animal, does owner have a farmer or not, annual animal leather, the farm area/fadan and animal contributes in daily living requirements or not. Suitable interview times were arranged. The sample size 758 we use simple random sample (SRS) for collected data and we use the method of principal components analysis to extract the important factors, which explain the variance.

Results and discussion

Descriptive statistics shows that seven hundred and fifty eight respondents were participated in this study. The majority of the respondents (93%) were males , and (7%) were females and twenty eight percent of the animal grazing hours/day (4 – 8) hours and seventy two percent over 8 hours and Seventy five percent animal vaccinated, twenty five percent not vaccinated, also ninety six percent of animal contributed in daily living requirements against four percent not contribute. Livestock water resources were distributed as follows: (67.9%) Boreholes, (19.8%) Stream – dam and (12.3%) more than one source, the education levels of the respondents: the majority (71%) were literacy, (21.2%) basic education, (5%) secondary schools and (2.8%) university education. The results of the factor analysis for the data of animal resources in north Kordofan state showed the underlying structure in table1. Twelve factors are extracted, which explain 63.12 of the total variance. The most important three factors are the health of live stock, farm area and experience of owner. The first factor which called (Animal Health) includes the variables (Livestock, Water Resource, Vaccinated Animal), which explains 8.25% about of the total variance. The second factor called (Range) includes the variables (Owner Farm, Farm Area), which explains 6.6% about of the total variance. The third factor called (Type of Owner) includes the variables (Owner Main Professional, Owner Gender), which explains 6.2% about of the total variance. The fourth factor called (Slaughter & Breeding) includes the variables (female Slaughtering Animal per Year, Breeding Type and Animal Commercial Domestic), which explains 5.8% about of the total variance. The fifth factor called (Animal Husbandry) includes the variables (Breeding Farm Area, Animal Health), which explains 5.25% about of the total variance. The sixth factor called (Animal Export) includes the variables (Export Animal, Owner Education Level), which explains 5.12% about of the total variance. The seventh factor called (Labor & Leather) includes the variables (Animal Labor Personnel, Annual Animal leather), which explains 5% about of the total variance. The eights factor called (Animal Mortality) includes the variables (Vaccinated

Extraction Method: Principal Component Analysis.

Animals, Reasons for Animal Decrement), which explains 4.61% about of the total variance. The ninth factor called (Leathers) includes the variables (lost of animal, grazing hours/day), which explains 4.28% about of the total variance. The tenth factor called (Animal Safety) includes the variables (Owner Professional Experience, Vaccinated Animals), Male Animal/Year Owner Professional Experience), which explains 3.8% about of the total variance. which explains 4.22% about of the total variance. The eleventh factor includes the variables (Owner Professional Experience and Animal Commercial Domestic), which explains 3.9% about of the total variance.

The twelve factor called (Animal Care) includes the variables (Slaughtering

Totbl 1: shows the Eigen-values associated with liner component (factor) before extraction

Component	Initial Eigen-values		
	Total	% of Variance	Cumulative %
1	2.144	8.245	8.245
2	1.720	6.614	14.860
3	1.611	6.197	21.057
4	1.518	5.839	26.896
5	1.364	5.248	32.144
6	1.328	5.107	37.251
7	1.305	5.018	42.269
8	1.197	4.606	46.874
9	1.112	4.278	51.153
10	1.096	4.216	55.369
11	1.017	3.910	59.279
12	1.003	3.859	63.138
13	.938	3.608	66.746
14	.916	3.524	70.270
15	.882	3.393	73.662
16	.876	3.368	77.030
17	.782	3.008	80.039
18	.755	2.903	82.942
19	.728	2.801	85.742
20	.702	2.702	88.444
21	.690	2.654	91.098
22	.576	2.216	93.314
23	.553	2.128	95.442
24	.497	1.911	97.353
25	.427	1.643	98.997
26	.261	1.003	100.000

According to Kaiser Criterion only the factors with eigenvalues greater than one is extracted.

Journal of Natural and Medical Sciences (JNMS)
vol. **15** (2) 2014
ISSN 1858-6805
e-ISSN 1858-6813

Table 2: Rotation Sums of Squared Loadings

Component	Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %
1	1.805	6.944	6.944
2	1.573	6.048	12.992
3	1.539	5.919	18.911
4	1.499	5.767	24.678
5	1.498	5.762	30.440
6	1.349	5.187	35.628
7	1.301	5.005	40.633
8	1.241	4.774	45.406
9	1.194	4.591	49.998
10	1.157	4.450	54.448
11	1.150	4.424	58.872
12	1.109	4.266	63.138

In table2 the eigenvalues associated with the twelve factors and the percentages of variance explained.

CorrelationMatrix

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}	x_{17}	x_{18}	x_{19}	x_{20}	x_{21}	x_{22}	x_{23}	x_{24}	x_{25}	x_{26}
x1		.000	.019	.071	.001	.000	.000	.036	.489	.103	.015	.338	.017	.179	.081	.193	.250	.000	.000	.000	.000	.089	.072	.099	.480	.401
x2	.000		.170	.113	.070	.268	.020	.312	.029	.371	.238	.143	.004	.030	.374	.119	.437	.064	.135	.061	.088	.140	.000	.015	.049	.161
x3	.019	.170		.264	.186	.001	.130	.374	.348	.052	.153	.379	.429	.315	.006	.469	.016	.395	.189	.040	.218	.118	.282	.102	.127	.020
x4	.071	.113	.264		.192	.074	.003	.400	.256	.377	.109	.376	.000	.410	.026	.005	.268	.190	.047	.188	.001	.151	.364	.206	.174	.010
x5	.001	.070	.186	.192		.031	.399	.129	.277	.067	.331	.407	.378	.004	.283	.361	.412	.008	.000	.409	.006	.000	.323	.374	.031	.129
x6	.000	.268	.001	.074	.031		.006	.065	.060	.000	.327	.249	.338	.069	.000	.042	.009	.464	.231	.015	.004	.001	.010	.041	.319	.088
x7	.000	.020	.130	.003	.399	.006		.013	.274	.210	.032	.208	.000	.340	.061	.117	.338	.000	.422	.468	.024	.122	.246	.098	.058	.016
x8	.036	.312	.374	.400	.129	.065	.013		.305	.165	.006	.011	.001	.087	.000	.008	.094	.127	.459	.322	.458	.046	.082	.156	.297	.237
x9	.489	.029	.348	.256	.277	.060	.274	.305		.288	.225	.371	.053	.201	.448	.047	.016	.274	.303	.000	.233	.203	.286	.117	.456	.068
x10	.103	.371	.052	.377	.067	.000	.210	.165	.288		.342	.443	.181	.381	.083	.011	.358	.029	.002	.138	.151	.180	.080	.000	.405	.482
x11	.015	.238	.153	.109	.331	.327	.032	.006	.225	.342		.283	.074	.493	.062	.021	.053	.000	.292	.001	.431	.347	.378	.475	.498	.207
x12	.338	.143	.379	.376	.407	.249	.208	.011	.371	.443	.283		.413	.249	.411	.481	.338	.070	.255	.310	.046	.014	.196	.103	.192	.357
x13	.017	.004	.429	.000	.378	.338	.000	.001	.053	.181	.074	.413		.484	.235	.000	.058	.050	.294	.000	.044	.068	.029	.000	.053	.456
x14	.179	.030	.315	.410	.004	.069	.340	.087	.201	.381	.493	.249	.484		.209	.157	.000	.363	.478	.447	.102	.252	.003	.042	.001	.399
x15	.081	.374	.006	.026	.283	.000	.061	.000	.448	.083	.062	.411	.235	.209		.298	.000	.496	.005	.019	.133	.418	.164	.004	.005	.478
x16	.193	.119	.469	.005	.361	.042	.117	.008	.047	.011	.021	.481	.000	.157	.298		.500	.035	.143	.480	.005	.051	.336	.004	.276	.037
x17	.250	.437	.016	.268	.412	.009	.338	.094	.016	.358	.053	.338	.058	.000	.000	.500		.150	.199	.147	.003	.032	.069	.320	.000	.112

Journal of Natural and Medical Sciences (JNMS)
vol. **15** (2) 2014
ISSN 1858-6805
e-ISSN 1858-6813

x18	.000	.064	.395	.190	.008	.464	.000	.127	.274	.029	.000	.070	.050	.363	.496	.035	.150		.304	.000	.181	.311	.301	.001	.096	.221
x19	.000	.135	.189	.047	.000	.231	.422	.459	.303	.002	.292	.255	.294	.478	.005	.143	.199	.304		.109	.094	.071	.045	.231	.132	.041
x20	.000	.061	.040	.188	.409	.015	.468	.322	.000	.138	.001	.310	.000	.447	.019	.480	.147	.000	.109		.013	.005	.009	.003	.325	.031
x21	.000	.088	.218	.001	.006	.004	.024	.458	.233	.151	.431	.046	.044	.102	.133	.005	.003	.181	.094	.013		.000	.242	.000	.078	.006
x22	.089	.140	.118	.151	.000	.001	.122	.046	.203	.180	.347	.014	.068	.252	.418	.051	.032	.311	.071	.005	.000		.241	.005	.294	.018
x23	.072	.000	.282	.364	.323	.010	.246	.082	.286	.080	.378	.196	.029	.003	.164	.336	.069	.301	.045	.009	.242	.241		.141	.000	.138
x24	.099	.015	.102	.206	.374	.041	.098	.156	.117	.000	.475	.103	.000	.042	.004	.004	.320	.001	.231	.003	.000	.005	.141		.268	.418
x25	.480	.049	.127	.174	.031	.319	.058	.297	.456	.405	.498	.192	.053	.001	.005	.276	.000	.096	.132	.325	.078	.294	.000	.268		.180
x26	.401	.161	.020	.010	.129	.088	.016	.237	.068	.482	.207	.357	.456	.399	.478	.037	.112	.221	.041	.031	.006	.018	.138	.418	.180	

Discussion:

Principal components analysis requires that there must be some correlations greater than 0.30 between the variables included in the analysis. For this set of variables, there are over 70 correlations in the matrix greater than 0.30, satisfying this requirement.

Table 3:The KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.518
Bartlett's Test of Sphericity	Approx. Chi-Square	1.911E3
	df	325
	Sig.	.000

The source: SPSS output

Table3 shows the Kaiser – Meyer – Olkin measure of sampling adequacy and Bartletts test of sphericity (Bartlett (1954)[9] (which evaluates the null hypothesis that all of the correlations in a correlation matrix are zero), SPSS obtained the result $\chi^2_{(325)} = 1911$, $p = .000$ (which indicates a probability less than .0005) for Barrtletts test of sphericity. Note that, χ^2 calculated is greater than χ^2 tabulated; the null hypothesis can be rejected at the 0.01

and 0.05 level. The value of KMO is 0.518, which indicates that the samples size is mediocre, so the sample size is sufficient. The Barrtletts test P- value is highly significant (0.000), thus it tells us that the correlation matrix is not identity matrix; therefore, the underlying population from which the sample was derived.

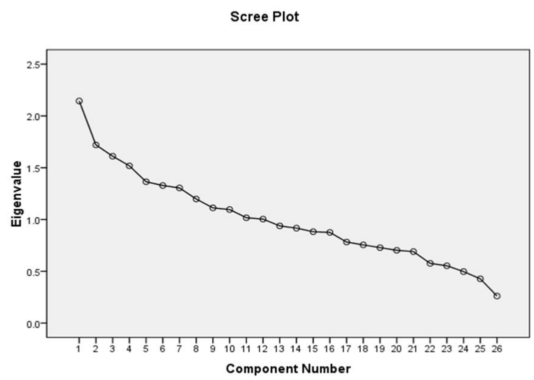

Figure 1: principal components and eigen values

For our livestock data, twelve components have eigenvalues greater than 1. On figure1, together components 1 to 12 account for 63% of the total variance. We shall retain only twelve components.

CONCLUSIONS

The study concluded that the more important factors increase livestock population water design livestock campaign and provision water resources and in decrease was export female of live stock.

REFERENCES

1- Rabie Ragab Sadek (1984) *Studies on genetic and environmental factor affecting the milk production of the Egyptian buffalo*, thesis of Doctor in philosophy on agriculture sciences, Department of animal production, faculty of agriculture, Cairo University Egypt.

2- Ahmed Hassan Mohamed Haridy (2010*). Studies on some factors affecting acquired immunity in chickens*, thesis of Doctor in philosophy on agriculture sciences (poultry sciences) Department of animal production, faculty of agriculture, Cairo University Egypt

3- Hair, J.F. & Black,W.C,(2000). Culster analysis. In Grimm, L.G. & Yarnold, P.R. (Eds), *Reading and understanding more multivariate statistics* (pp. 147-205). Washington, DC: American Psychological Association.

4- Darlington, Richard B., Sharon Weinberg, and Herbert Walberg (1973). *Canonical variate analysis and related techniques*. Review of Educational Research, 453-454.

5- Chacon E A, Stobbs T H and Dale M B. (1978). *Influence of sward characteristics on grazing behaviour and growth of Hereford steers grazing tropical grass pastures*. *Australian Journal of Agricultural Research* 29(1):89-102.

6- *grazing time of cattle in a subhumid African savanna*. In: Tothill J C and Mott J J (eds) Ecology and management of the world's savannas. Australian Academy of Science, Canberra, Australia. pp. 256-259.

7- King J M. (1983*). Livestock water needs in pastoral Africa in relation to climate and forage. ILCA Research Report 7*. ILCA (International

Livestock Centre for Africa), Addis Ababa, Ethiopia. 95 pp.

8- Jonathon Shlens *A Tutorial on Principal Component Analysis* http://www.brainmapping.org/NITP/PNA/Readings/pca.pdf 2005/12/10

9- Dochtermann, N.A. & Jenkins S.H. 2007: *Behavioural syndromes in Merriam's kangaroo rats (Dipodomys merriami): A test of competing hypotheses. Proc. R. Soc*. London B 274, 2343-2349.