

Data analysis methods adopted under *in situ* global project in Nepal

Bal K Joshi, Madhusudan P Upadhyay, Hari P Bimb, D Gauchan and BK Baniya

Nepal Agricultural Research Council, PO Box 1135 Kathmandu-Nepal <joshibalak@rdiffmail.com>

ABSTRACT

Synthesizing data analysis methods adopted under *in situ* global project in Nepal along with variables and nature of study could be guiding reference for researchers especially to those involved in on farm research. The review work was conducted with the objective to help in utilizing and managing *in situ* database system. The objectives of the experiment, the structure of the treatments and the experimental design used primarily determine the type of analysis. There were 60 papers of this project published in Nepal. All these papers are grouped under 8 thematic groups namely 1. Agroecosystem (3 papers), 2. Agromorphological and farmers' perception (7 papers), 3. Crop population structure (5 papers), 4. Gender, policy and general (15 papers), 5. Isozyme and molecular (6 papers), 6. Seed systems and farmers' networks (5 papers), 7. Social, cultural and economical (11 papers) and 8. Value addition (8 papers). All these papers were reviewed basically for data type, sample size, sampling methods, statistical methods and tools, varieties and purposes. Descriptive and inferential statistics along with multivariate methods were commonly used in on farm research. Experimental design, the most common in on station trial was least used. Study over space and time was not adopted. There were 5 kinds of data generated, 45 statistical tools adopted in eight different crop species. Among the 5 kinds of data under these eight subject areas, categorical type was highest followed by discrete numerical. Binary type was least in frequency. Most of the papers were related to rice followed by taro and finger millet. Cucumber and pigeon pea were studied least. Descriptive statistics along with χ^2 , multivariate analysis and regression approaches would be appropriate tools. Similarly SPSS and MINITAB may be good software. The best one among a number of statistical tools should be selected and utmost care must be exercised while collecting data.

Key words: Data analysis methods, on farm research, on station research, subject areas

INTRODUCTION

Several statistical methods have been developed, tested and routinely used in on station research system. Analysis of variance is the most common way of data management among NARC scientists (Joshi and Shrestha 2003). Limited literatures are available for analyzing and interpreting the on farm and *in situ* research data. In 1995, International Plant Genetic Resources Institute (IPGRI) together with national partners from nine countries covering five regions, Sub Saharan Africa, Central and West Asia and North Africa, Asia the Pacific and Oceania, the Americas and Europe formulated a global project to strengthen the scientific basis of *in situ* conservation of agricultural biodiversity on farm. Achievements made on the method of data handling from this project are valued among researchers especially those involved in on farm trials (Jarvis et al. 2000, Jarvis and Hodgkin 1998). The project research team have used many statistical tools in different subject areas and interpreted accordingly. Synthesizing this method along with variables and nature of study could be guiding reference for researchers especially to those involved in on farm research. There are many statistical methods whose purpose is to help understand data. Care should be taken during walking in the road from the experimentation to publication. A bad design implemented in field would result in the death of the experiment. If however, an experiment is well designed and executed, a subsequent bad analysis can be corrected. The objectives of the experiment, the structure of the treatments and the experimental design used primarily determine the type of analysis. In many journals authors (Baker 1980, Carmer and Swanson 1971, Carmer 1976, Carmer et al 1979, Joshi et al 2002) pointed out the errors in the use of statistical techniques. Utilization and management of *in situ* data base system is necessary for developing *in situ* conservation methods and models. Therefore the methods and models of data handling under *in situ* global project-Nepal component was assessed. Basically we have synthesized

and discussed statistical tools, sample size, variables, software used, and nature of study of on farm research in Nepal. Besides additional statistical tools are given with respect to subject area of study.

METHODOLOGY

All 60 papers of in situ global project-Nepal component (1999-2000) were reviewed and interpreted data analysis methods. All these papers are grouped under 8 thematic areas ie 1. Agroecosystem (3 papers), 2. Agromorphological and farmers' perception (7 papers), 3. Crop population structure (5 papers), 4. Gender, policy and general (15 papers), 5. Isozyme and molecular (6 papers), 6. Seed systems and farmers' networks (5 papers), 7. Social, cultural and economical (11 papers) and 8. Value addition (8 papers). Three site selection reports (Sherchand et al 1998, Poudyal et al 1998, Rijal et al 1998) and three baseline survey reports (Rana et al 2000a, 2000b, 2000c) were included under social, cultural and economical group. All other papers were from two proceedings ie Sthapit et al (2003) and Sthapit et al (2000). All these papers were reviewed basically for data type, sample size, sampling methods, statistical methods and tools, varieties and purposes. Frequency was used to interpret the findings. Statistical tools were analyzed in depth and appropriate tools and interpretation are suggested. A training manual written by Jarvis et al (2000) and Mutsaers et al (1997) were consulted for listing additional potential statistical tools.

RESULTS AND DISCUSSION

Data collecting, processing and analyzing methods along with purposes and nature of study are given in Table 1. There were some differences on data handling methods among subject area of study. Descriptive statistics along with multivariate methods were the common in on farm research and experimental design which is most common in on station trial was least used. Study over space and time was not adopted. With respect to subject area, frequency of data type collected, statistical tools used and crop studied are given in Table 2. There were 5 kinds of data generated, 45 statistical means adopted in eight different crop species. These methods and crops were used for answering 4 research questions, i. What is the amount and distribution of genetic diversity maintained by farmers over space and time, ii. What are the processes used to maintain genetic diversity on farm, iii. Who maintain diversity on farm and iv. What are the factors that influence farmer's decision making to maintain diversity. In addition to these tools used by on farm researchers, there are many other tools that help to interpret data. Some of them are listed in Annex 1. Farmers are major source for data collection in on farm trials. Therefore, farmers must be cooperative and clear about objectives of the study.

Figure 1. Total use percentage of data types (A) and total times of crop species (B) considered in on farm research study in 60 papers.

**T
a
b
l
e
2
.
F
r
e
q
u
e
n
c
y
o
f
d
a
t
a
t
y
p
e
a
n
d
s
t
a
t
i
s
t
i
c
a
l
t
o
o
l
s
u
s
e
d
i
n
6
0
p
a
p**

Variables	Sectors															Total	
	AE		AFP		CP		GP		IM		SS		SCE		VA		
	fi	%	fi	%	fi	%	fi	%	fi	%	fi	%	fi	%	fi		%
A. Data type																	
Binary	-	-	-	-	-	-	1	6.67	6	100	1	20	1	9.09	-	-	9
Categorical	3	100	7	100	5	100	2	13.3	-	-	5	100	11	100	2	25	35
Numerical (continuous)	2	66.7	6	85.7	5	100	2	13.3	-	-	-	-	11	100	3	37.5	29
Numerical (discrete)	3	100	6	85.7	5	100	4	26.7	-	-	1	20	10	90.9	1	12.5	30
Perception	-	-	-	-	-	-	10	66.7	-	-	-	-	-	-	3	37.5	13
Total papers	3	-	7	-	5	-	15	-	6	-	5	-	11	-	8	-	60
B. Statistical tools																	
%	2	66.7	5	71.4	4	80	2	13.3	5	83.3	5	100	10	90.9	4	50	37
A	-	-	-	-	-	-	-	-	4	66.7	-	-	-	-	-	-	4
ANOVA	-	-	-	-	1	20	-	-	-	-	-	-	4	36.4	-	-	5
Ap	-	-	-	-	-	-	-	-	4	66.7	-	-	-	-	-	-	4
χ^2	1	33.3	-	-	-	-	-	-	-	-	-	-	4	36.4	-	-	5
Cluster	-	-	5	71.4	2	40	-	-	6	100	-	-	3	27.3	-	-	16
Coding	-	-	1	14.3	-	-	-	-	-	-	-	-	-	-	-	-	1
CV	-	-	4	57.1	4	80	-	-	-	-	-	-	-	-	1	12.5	9
D	-	-	-	-	-	-	1	6.67	-	-	-	-	-	-	-	-	1
Dice coefficient	-	-	-	-	-	-	-	-	1	16.7	-	-	-	-	-	-	1
Distribution	1	33.3	3	42.9	1	20	1	6.67	-	-	-	-	1	9.09	2	25	9

Variables	Sectors															Total	
	AE		AFP		CP		GP		IM		SS		SCE		VA		
	fi	%	fi	%	fi	%	fi	%	fi	%	fi	%	fi	%	fi		%
Dst	-	-	-	-	-	-	-	-	3	50	-	-	-	-	-	-	3
fi	1	33.3	4	57.1	2	40	1	6.67	1	16.7	4	80	2	18.2	3	37.5	18
Genetic erosion	-	-	-	-	-	-	1	6.67	-	-	-	-	2	18.2	-	-	3
Goodness of fit	-	-	-	-	-	-	-	-	1	16.7	-	-	-	-	-	-	1
Gst	-	-	-	-	-	-	-	-	3	50	-	-	-	-	-	-	3
H'	-	-	1	14.3	-	-	1	6.67	-	-	-	-	-	-	-	-	1
H	-	-	-	-	-	-	-	-	3	50	-	-	-	-	-	-	4
Hs	-	-	-	-	-	-	-	-	3	50	-	-	-	-	-	-	3
Ht	-	-	-	-	-	-	-	-	3	50	-	-	-	-	-	-	3
LSD	-	-	-	-	1	20	-	-	-	-	-	-	-	-	-	-	1
Mental test	-	-	-	-	-	-	-	-	1	16.7	-	-	-	-	-	-	1
Matrix ranking	-	-	-	-	-	-	-	-	-	-	-	-	3	27.3	1	12.5	4
Maximum	-	-	-	-	-	-	-	-	-	-	-	-	3	27.3	-	-	3
Mean	2	66.7	7	100	5	100	2	13.3	6	100	-	-	8	72.7	2	25	32
Median	-	-	-	-	-	-	-	-	-	-	-	-	1	9.09	-	-	1
Minimum	-	-	-	-	-	-	-	-	-	-	-	-	3	27.3	-	-	3
Organoleptic test	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	12.5	1
P	-	-	-	-	-	-	-	-	4	66.7	-	-	-	-	-	-	4
Pair-t test	-	-	-	-	-	-	-	-	-	-	-	-	2	18.2	-	-	2
PCA	-	-	4	57.1	2	40	-	-	5	83.3	-	-	-	-	-	-	11
PIC	-	-	-	-	-	-	-	-	1	16.7	-	-	-	-	-	-	1
Proportion	-	-	1	14.3	-	-	1	6.67	-	-	-	-	-	-	-	-	2
r	1	33.3	-	-	-	-	-	-	3	50	-	-	3	27.3	1	12.5	8
Range	-	-	5	71.4	4	80	-	-	3	50	-	-	1	9.09	3	37.5	16
Ranking	-	-	-	-	-	-	-	-	-	-	-	-	2	18.2	-	-	2
Rating	1	33.3	-	-	-	-	-	-	-	-	-	-	3	27.3	-	-	4
Ratio	-	-	-	-	1	20	1	6.67	-	-	-	-	2	18.2	1	12.5	5
Richness	2	66.7	2	28.6	1	20	2	13.3	-	-	2	40	6	54.5	-	-	15
Scoring	1	33.3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
SD	-	-	6	85.7	3	60	1	6.67	3	50	-	-	6	54.5	2	25	21
SE	-	-	-	-	3	60	-	-	1	16.7	-	-	-	-	1	12.5	5
Trend analysis	-	-	-	-	-	-	1	6.67	-	-	1	20	-	-	1	12.5	3
Tukey	-	-	-	-	-	-	-	-	1	16.7	-	-	-	-	-	-	1
Zymotype	-	-	-	-	-	-	-	-	3	50	-	-	-	-	-	-	3
Total papers	3	-	7	-	5	-	15	-	6	-	5	-	11	-	8	-	60
C. Crops																	
Barley	2	66.7	1	14.3	2	40	-	-	2	33.3	-	-	1	9.09	-	-	8
Buckwheat	2	66.7	-	-	2	40	-	-	2	33.3	-	-	1	9.09	1	12.5	8
Cucumber	2	66.7	-	-	1	20	-	-	-	-	-	-	3	27.3	-	-	6
Finger millet	2	66.7	1	14.3	1	20	2	13.3	-	-	1	20	3	27.3	1	12.5	11
Pigeon pea	2	66.7	1	14.3	1	20	1	6.67	-	-	-	-	1	9.09	-	-	6
Rice	2	66.7	2	28.6	3	60	4	26.7	1	16.7	3	60	8	72.7	6	75	29
Sponge gourd	2	66.7	3	42.9	-	-	-	-	-	-	-	-	2	18.2	-	-	7
Taro	3	100	2	28.6	1	20	1	6.67	1	16.7	1	20	1	9.09	1	12.5	11
Not crop specific	-	-	-	-	-	-	10	66.7	-	-	-	-	3	27.3	2	25	15
Total papers	3	-	7	-	5	-	15	-	6	-	5	-	11	-	8	-	60

AE, Agro-ecosystem. AFP, Agromorphological and farmers' perception. CP, Crop population. GP, Gender, policy and general. IM, Isozyme and molecular. SS, Seed supply and farmers' network. SCE, Social, cultural and economical. VA, Value addition. %, Percent. ANOVA, Analysis of variance, CRD, Completely randomized design, CV, Coefficient of variance, D, Simpson index, FGD, Focus group discussion, fi, Frequency, GIS, Geographic

information system, Hb, Brillion index, H, Gene diversity, H', Shannon weaver index, HHS, Household survey, HH, Household, KI, Key informant, KII, Key informant interview, LRS, Literature survey, LSD, Lest significant difference, NA, Not applicable, N_e , Nei's index, PCA, Principal components analysis, PIC, Polymorphic information index, PRA, Participatory rural appraisal, r, Correlation coefficient, RCBD, Randomized complete block design, SD, Standard deviation, SE, Standard error, N, Average number of individuals sampled (mean across all loci). L, Nnumber of loci sampled. A, Mean number of alleles per locus. Ap, Mean number of alleles per polymorphic locus. P, Proportion of polymorphic loci. H, Expected heterozygosity. This value will tend to be lower than the actual value if the number of individuals sampled is low. U, Number of alleles unique to that population (=private alleles). $p(1)$, Average frequency of private alleles, useful for obtaining indirect estimates of gene flow. Ht, Total gene diversity. Hs, Mean gene diversity within population. Dst, Mean gene diversity among populations. Gst, Coefficient of gene differentiation. Hs/Ht, Percentage mean gene diversity within population.

Agro- ecosystem

Researchers under this subject have tried to explore, characterize, evaluate and to know the distribution pattern considering categorical and numerical types of variables of crop species and households. Survey was the major tool to generate data. On farm and on station methods were also used to generate data. Among the 9 statistical tools, most common were percentage, mean and richness. Eight crop species were used. Diversity (alpha, beta, gama) and classification methods are additional appropriate tools in such type of study. Spatial and temporal study should also be considered.

Agromorphological and farmer perception

There were 7 papers under this category. Survey, on farm and on station experiments were the methods of generating data with objectives of locating and measuring diversity and characterizing variables. On farm method with observatory type was common and MINITAB was used by most of the scientists. Sample was chosen randomly with great variation in sample size. Buckwheat and cucumber were not agromorphologically studied. Categorical type of data was highest in frequency. Among statistical tools mean was most commonly used followed by standard deviation (SD). Three papers were related to sponge gourd, which is highest in number among six crop species. Multivariate methods were also common in this subject area but their interpretation and presentation style were not well and impressive. Nature of study and set objective should be considered during choosing statistical tools. χ^2 and pattern analysis may be good methods for interpreting agromorphological and farmers' perception data.

Crop population

Objectives under this category were related to population structure, diversity and characterization. Sample was chosen randomly in most of the cases. Common experimental method was on farm with observation type, which was handled mostly by Excel software. Three variables were used in 7 crop species. Mean was the most common tools followed by percentage, coefficient of variance (CV), and range. Rice had received the highest priority for population structure study. Here principle component analysis (PCA) and cluster were not used property. Distribution at ecosystem, species and genetic levels will be core area of study in this category.

Gender, policy and general

The highest number of papers was related to gender and policy. Samples were selected purposively and data were generated mostly from survey. Perception type of data was highest in frequency followed by discrete numerical type. Percentage, mean and richness were commonly used tools. Most of the papers were not relevant to crop species. In addition to these tools, relationship, 4-square analysis and ordination methods may be used. Least number of statistical tools has been used in this subject area.

Isozyme and molecular

Only binary data were used with the objectives of evaluation, verification, diversity and genetic relationship in four crop species. This is totally lab-based research and many genetic parameters were estimated. Most of these binary data was handled by NTSys software. Use of cluster and mean were common. Unique tools eg Mental test and Cophenetic correlation etc have been used. Major issue raised

here is the application of these findings. Spatial and temporal analysis along with distance measures may be other tools that can help to draw inference in such a study.

Seed supply and farmer's network

Households were the main variables studied to know the seed supply and storage system along with diversity maintenance. Most of the samples were purposively selected and all authors have generated data through survey. Categorical data type was highest in frequency. Percentage and frequency were frequently used. Study was done in only 3 crop species, finger millet, rice and taro.

Social, cultural and economical

More details information was drawn under this subject area by 11 authors in all mandated crops. There were many variables taken to characterize, evaluate, compare, relate and to assess diversity. Randomly sampled information was generated through survey in most of the cases. Data were analyzed using SPSS. Categorical and continuous numerical data dominated in this research. Percentage was frequently used and second one was mean. Rice was relatively studied many more times than other crop species. There was a use of PCA, cluster and analysis of variance (ANOVA) but their presentation and inference drawing were not impressive. Some of additional tools are 4-square analysis, ordination, classification methods, regression and trend analysis that can help to draw more valid information.

Value addition

This aspect was studied in rice, taro, buckwheat, and finger millet. Survey was the common means of data generation from purposively selected sample with an objective of adding benefit. Perception and continuous numerical data were common which were mostly described by percentage. This study was concentrated more on rice.

Summary in use of statistical tools, data types and crop species on percentage are given in Figure 1, 2. Among the 5 types of data type under these eight subject areas, categorical type was highest followed by discrete numerical (Figure 1). Binary type was least in frequency. Most of the papers were related to rice followed by taro and finger millet. Cucumber and pigeon pea were studied least. Percentage was the most common tool and second most frequently used tool was mean. Genetic parameters were estimated by least number of authors (Figure 2). In addition to these tools, there are many tools. Appropriate tools should be selected based on the objective of study. In on farm research variance can't be controlled but we have to use properly during interpretation. Special design was not followed and farmer was used as a replicate. Village can be used as a block. In such cases, descriptive statistics along with χ^2 , multivariate analysis and regression approaches are appropriate. Similarly SPSS and MINITAB may be good software to handle such type of data. Excel is the excellent to enter data and to process them. There are needs of improvement on interpretation, presentation of findings and drawing of conclusions. A scientific method of research ie first identify problem, set objective, formulate hypothesis and test the validity should be followed. To validate the hypothesis, data should be collected properly, analyzed and presented in a good manner and in the last findings should be generalized. For interpretation of the results, there are a number of statistical tools those can help in getting logical inference. But researcher should be able to select best one and utmost care must be exercised while collecting data.

Most of the tools used for managing and utilizing on farm data were similar to on station trails. These were used basically to answer above-mentioned four research questions. This report will be useful for developing common on farm database systems and hypothesis testing.

ACKNOWLEDGEMENTS

We thank BR Sthapit and S Gyawali for their valuable comments.

REFERENCES

- A (Bajracharya et al 1999, Baniya et al 1999, Joshi et al 1999, Rijal et al 1999, Khatiwada et al 1999, Paudel et al 1999, Gauchan 1999, Upadhyay and Subedi 1999, Subedi et al 1999, Gauchan et al 1999). **In:** *A scientific basis of in situ conservation of agrobiodiversity on farm: Nepal's contribution to the global project* (B Sthapit, M Upadhyay and A Subedi, eds). NP Working Paper No. 1/99.
- B (Bajracharya et al 2003a, Bajracharya et al 2003b, Bimb et al 2003, Bajracharya et al 2003c, Bajracharya et al 2003d, Gupta et al 2003, Yadav et al 2003a, Rijal et al 2003b, Pandey et al 2003a, Yadav et al 2003b, Tiwari et al 2003a, Pandey et al 2003b, Baniya et al 2003a, Khatiwada et al 2003, Tiwari et al 2003b, Rana et al 2003b, Gauchan et al 2003, Chaudhary et al 2003c, Rijal et al 2003a, Joshi et al 2003a, Rijal et al 2003c, Joshi et al 2003b, Rijal et al 2003d, Pant et al 2003, Joshi et al 2003c, Joshi et al 2003d, Sthapit and Jarvis 2003, Upadhyay and Subedi 2003, Rana et al 2003a, Chaudhary et al 2003a, Upadhyay et al 2003, Sapkota et al 2003, Subedi et al 2003a, Chaudhary et al 2003b, Adhikari 2003, Sah and Sah 2003, Adhikari and Adhikari 2003, Thapa 2003, Subedi et al 2003b, Baniya et al 2003b, Baniya et al 2003c, Baniya et al 2003d). **In:** *On farm management of agricultural biodiversity in Nepal* (BR Sthapit, MP Upadhyay, BK Baniya, A Subedi and BK Joshi, eds). Proceedings of a National workshop, 24-26 April 2001 Lumle Nepal. NARC, LIBIRD and IPGRI.
- Baker, RJ. 1980. Multiple comparison tests. *Can. J. Plant Sci.* 60:325-327.
- Carmer SG. 1976. Optimal significance levels for application of the least significant difference in crop performance trials. *Crop Sci.* 16:95-99.
- Carmer SG and MR Swanson. 1971. Detection of differences between means: A Monte Carlo study of five pair wise multiple comparison procedures. *Agron. J.* 63:940-945.
- Carmer SG, WM Walker and RD Seif. 1979. Statisticians as reviewers. *Hort. Science* 11:348-357.
- Jarvis DI, L Myer, H Klemick, L Guarino, M Smale, AHD Brown, M Sadiki, B Sthapit and T Hodgkin. 2000. *A training guide for in situ conservation on farm*. Version 1. IPGRI, Rome.
- Jarvis DI and T Hodgkin (eds). 1998. Strengthening the scientific basis of *in situ* conservation of agricultural biodiversity on farm: Options for data collecting and analysis. *Proceedings of a workshop to develop tools and procedures for in situ conservation on farm*, 25-29 Aug 1997. IPGRI Rome.
- Joshi BK, S Gyawali and DS Poudyal. 2002. Regression analyses and multiple comparison procedures: Uses and misuses. *JIST.* 12:69-81.
- Joshi BK and NP Shrestha. 2003. Statistical software used by NARC scientists. *NARC Newsletter* 10(1):4-5.
- Mutsaers HJW, GK Weber, P Walker and NM Fisher. 1997. *A field guide for on farm experimentation*. IITA/CTA/ISNAR, Nigeria.
- Poudyal CL, PR Tiwari, JD Neupane and DP Devkota. 1998. *Findings of site selection in Jumla, Nepal*. NP Working Paper No. 3/98. NARC, LIBIRD and IPGRI.
- Rana RB, CL Paudyal, PR Tiwari, D Gauchan, A Subedi, BR Sthapit, MP Upadhyay and DI Jarvis. 2000J. *In situ crop conservation: Findings of agro-ecological, crop diversity and socio-economic baseline survey of Talium ecosite, Jumla, Nepal*. NP Working Paper No 3/2000. NARC, LIBIRD and IPGRI.
- Rana RB, DK Rijal, D Gauchan, BR Sthapit, A Subedi, MP Upadhyay, YR Pandey and DI Jarvis. 2000K. *In situ crop conservation: Findings of agro-ecological, crop diversity and socio-economic baseline survey of Begnas ecosite, Kaski, Nepal*. NP Working Paper No. 2/2000. NARC, LIBIRD and IPGRI.
- Rana RB, P Chaudhary, D Gauchan, SP Khatiwada, BR Sthapit, A Subedi, MP Upadhyay and DI Jarvis. 2000B. *In situ crop conservation: Findings of agro-ecological, crop diversity and socio-economic baseline survey of Kachorwa ecosite, Bara, Nepal*. NP Working Paper No. 1/2000. NARC, LIBIRD and IPGRI.
- Rijal DK, RB Rana, KK Sherchand, BR Sthapit, YR Pandey, N Adhikari, KB Kadayat, YP Gautam, P Chaudhary, CL Poudyal, SR Gupta and PR Tiwari. 1998. *Findings of site selection in Kaski, Nepal*. NP Working Paper No. 1/98. NARC, LIBIRD and IPGRI.
- Sherchand KK, NP Adhikari, SP Khatiwada, AC Shrivastav, J Bajracharya, KD Joshi, KB Kadayat, M Chaudhary, P Chaudhary, SS Vishwakarma and S Yadav. 1998. *Strengthening the scientific basis for in situ conservation of agrobiodiversity: Findings of site selection in Bara, Nepal*. NP Working Paper No. 2/98. NARC, LIBIRD and IPGRI.

Annex 1. Additional statistical tools that can be used for on farm research (Refer Table 2 for full description)**Agro-ecosystem**

Classification method
 Hierarchical cluster analysis
 Non-hierarchical cluster analysis
 Modeclus cluster analysis
 Ordination Method (D2)
 Polar ordination
 Principal coordinate analysis
 Reciprocal averaging
 Detrended correspondence analysis
 Canonical variate analysis
 Canonical discriminant analysis
 Depended vs independent variables
 Multiple regression
 Canonical correlation
 Binary discriminant
 Multiple discriminant
 Discriminant analysis
 Relationship
 GIS
 Spatial and temporal analysis
 4-square analysis

Agromorphological
 Ranking
 GIS
 Univariate data analysis
 Mode
 Median
 ANOVA
 Minimum
 Maximum
 Diversity Indices
 Evenness
 D
 Hb
 Non parametric test
 Covariate analysis
 Bivariate Data analysis
 χ^2
 Similarity measures
 Simple matching coefficient
 Jaccard's coefficient
 Multivariate data analysis
 Classification method
 Non-hierarchical cluster
 analysis
 Ordination Method
 Polar ordination
 Principal coordinate analysis
 Reciprocal averaging
 Detrended correspondence
 analysis
 Canonical variate analysis
 Canonical discriminant analysis
 Factor analysis of correspondence
 Pattern analysis
 Distance measures
 Euclidean distance
 Mahalanobi's generalized distance
 Roger's distance
 Gower's similarity coefficient

Analysis of relationship among characters, individuals, populations, sites, sampling time, between results obtained from different sets of characters

Partitioning of variations within and between populations, sites, sampling times

4-square analysis

Crop population

Evenness
 Density
 Minimum viable population
 Allele distributions
 Out crossing rates
 Gene flow
 Landrace occurrence patterns
 Genotype diversity index
 D
 Net index of gene diversity
 Population size
 Variance
 Distributions
 Common
 Rare
 Widespread
 Local

Genetic structure

Polymorphic gene
 Allelic richness
 Frequency of allelic variants
 Level of heterozygosity
 Number of multilocus genotypes
 Distinctiveness
 Degree of linkage disequilibrium
 Fixation index
 Coefficient of gene diversity

Relationship

4-square analysis

Biochemical and molecular markers

Allelic richness
 Population subdivisions
 Effective population size
 Similarity
 Distance
 Genetic map

Spatial and temporal (time series) analysis

4-square analysis

t test

Seed system

Seed flow
 Fuzzy analysis
 N_e
 Ranking
 Rating
 Sorting
 Matrix ranking
 Time series analysis
 Auto correlation coefficient
 Cross correlation
 Power spectral analysis
 4-square analysis