World Bank – ICAR funded
National Agricultural Higher Education Project
Centre for Advanced Agricultural Science and Technology (CAAST)
On
Genomics Assisted Crop Improvement and Management

Training Manual
Advances in Research Methodology for Social Sciences

August 31 – September 4, 2020

Division of Agricultural Economics
ICAR – Indian Agricultural Research Institute
New Delhi – 110012
www.nahep-caast.iari.res.in

NAHEP Sponsored

Short term training programme

On

Advances in Research Methodology for Social Sciences

Course Convenor

Alka Singh
Professor and Head
Division of Agricultural Economics
ICAR-Indian Agricultural Research Institute
Pusa Campus, Delhi – 110012
Email: asingh.eco@gmail.com
Phone No. 9871198527

Co-Convenors

R.N. Padaria
Professor
Division of Agricultural Extension
ICAR-Indian Agricultural Research Institute
New Delhi 110 012
E-mail: rabi64@gmail.com

R. R Burman
Principal Scientist
Division of Agricultural Extension
ICAR-Indian Agricultural Research Institute
New Delhi 110 012
E-mail: burman_extn@hotmail.com

Aditya K.S.
Scientist
Division of Agricultural Economics
ICAR-Indian Agricultural Research Institute
New Delhi 110 012
E-mail: adityaag68@gmail.com

Praveen K.V.
Scientist
Division of Agricultural Economics
ICAR-Indian Agricultural Research Institute
New Delhi 110 012
E-mail: veenkv@gmail.com

Division of Agriculture Economics
ICAR-Indian Agricultural Research Institute
New Delhi- 110 012

About NAHEP-CAAST at IARI, New Delhi

Centre for Advanced Agricultural Science and Technology (CAAST) is a new initiative and student centric subcomponent of World Bank sponsored **National Agricultural Higher Education Project (NAHEP)** granted to the Indian Council of Agricultural Research, New Delhi to provide a platform for strengthening educational and research activities of post graduate and doctoral students. The ICAR-Indian Agricultural Research Institute, New Delhi was selected by the NAHEP-CAAST programme. NAHEP sanctioned Rs 19.99 crores for the project on "**Genomic assisted crop improvement and management**" under CAAST programme. The project at IARI specifically aims at inculcating genomics education and skills among the students and enhancing the expertise of the faculty of IARI in the area of genomics.

Objectives:

1. **To develop online teaching facility and online courses for enhancing the teaching and learning efficiency, and scientific communication skills**
2. **To develop and/or strengthen state-of-the art next-generation genomics and phenomics facilities for producing quality PG and Ph.D.students**
3. **To develop collaborative research programmes with institutes of international repute and industries in the area of genomics and phenomics**
4. **To enhance the skills of faculty and PG students of IARI and NARES**
5. **To generate and analyze big data in genomics and phenomics of crops, microbes and pests for genomics augmentation of crop improvement and management**

IARI's CAAST project is unique as it aimed at providing funding and training support to the M.Sc. and Ph.D. students from different disciplines who are working in the area of genomics. It will organize lectures and training programmes, send IARI students for training at expert laboratories and research institutions abroad, and cover students from several disciplines. It will provide opportunities to the students and faculty to gain international exposure. Further, the project envisages developing a modern lab named as **Discovery Centre** that will serve as a common facility for students' research at IARI.

Core-Team Members:

S.No.	Name of the Faculty	Discipline	Institute
1.	Dr. Ashok K. Singh	Genetics	ICAR-IARI
2.	Dr. Vinod	Genetics	ICAR-IARI
3.	Dr. Gopala Krishnan S	Genetics	ICAR-IARI
4.	Dr. A. Kumar	Plant Pathology	ICAR-IARI
5.	Dr. T.K. Behera	Vegetable Science	ICAR-IARI
6.	Dr. R.N. Sahoo	Agricultural Physics	ICAR-IARI
7.	Dr. Alka Singh	Agricultural Economics	ICAR-IARI
8.	Dr. A.R. Rao	Bioinformatics	ICAR-IASRI
9.	Dr. R.C. Bhattacharya	Molecular Biology & Biotechnology	ICAR-NIPB
10.	Dr. K. Annapurna	Microbiology **Nodal officer, Grievance Redressal, CAAST**	ICAR-IARI
11.	Dr. R. Roy Burman	Agricultural Extension **Nodal officer, Equity Action Plan, CAAST**	ICAR-IARI
12.	Dr. K.M. Manjaiah	Soil Science & Agri. Chemistry **Nodal officer, CAAST**	ICAR-IARI
13.	Dr.Viswanathan Chinnusamy	Plant Physiology **PI, CAAST**	ICAR-IARI

Associate Team

S.No.	Name of the Faculty	Discipline	Institute
14.	Dr. Kumar Durgesh	Genetics	ICAR-IARI
15.	Dr. Ranjith K. Ellur	Genetics	ICAR-IARI
16.	Dr. N. Saini	Genetics	ICAR-IARI
17.	Dr. D. Vijay	Seed Science & Technology	ICAR-IARI
18.	Dr. Kishor Gaikwad	Molecular Biology & Biotechnology	ICAR-NIPB
19.	Dr. Mahesh Rao	Genetics	ICAR-NIPB
20.	Dr. Veena Gupta	Economic Botany	ICAR-NBPGR
21.	Dr. Era V. Malhotra	Molecular Biology & Biotechnology	ICAR-NBPGR
22.	Dr. Sudhir Kumar	Plant Physiology	ICAR-IARI
23.	Dr. Dhandapani R	Plant Physiology	ICAR-IARI
24.	Dr. Lekshmy S	Plant Physiology	ICAR-IARI
25.	Dr. Madan Pal	Plant Physiology	ICAR-IARI
26.	Dr. Shelly Praveen	Biochemistry	ICAR-IARI
27.	Dr. Suresh Kumar	Biochemistry	ICAR-IARI
28.	Dr. Ranjeet R. Kumar	Biochemistry	ICAR-IARI
29.	Dr. S.K. Singh	Fruits & Horticultural Technology	ICAR-IARI
30.	Dr. Manish Srivastava	Fruits & Horticultural Technology	ICAR-IARI
31.	Dr. Amit Kumar Goswami	Fruits & Horticulture Technology	ICAR-IARI
32.	Dr. Srawan Singh	Vegetable Science	ICAR-IARI
33.	Dr. Gograj S Jat	Vegetable Science	ICAR-IARI
34.	D. Praveen Kumar Singh	Vegetable Science	ICAR-IARI
35.	Dr. V.K. Baranwal	Plant Pathology	ICAR-IARI
36.	Dr. Deeba Kamil	Plant Pathology	ICAR-IARI
37.	Dr. Vaibhav K. Singh	Plant Pathology	ICAR-IARI
38.	Dr. Uma Rao	Nematology	ICAR-IARI
39.	Dr. S. Subramanium	Entomology	ICAR-IARI
40.	Dr. M.K. Dhillon	Entomology	ICAR-IARI
41.	Dr. B. Ramakrishnan	Microbiology	ICAR-IARI
42.	Dr. V. Govindasamy	Microbiology	ICAR-IARI
43.	Dr. S.P. Datta	Soil Science & Agricultural Chemistry	ICAR-IARI
44.	Dr. R.N. Padaria	Agricultural Extension	ICAR-IARI
45.	Dr. Satyapriya	Agricultural Extension	ICAR-IARI
46.	Dr. Sudeep Marwaha	Computer Application	ICAR-IASRI
47.	Dr. Seema Jaggi	Agricultural Statistics	ICAR-IASRI
48.	Dr. Anindita Datta	Agricultural Statistics	ICAR-IASRI
49.	Dr. Soumen Pal	Computer Application	ICAR-IASRI
50.	Dr. Sanjeev Kumar	Bioinformatics	ICAR-IASRI
51.	Dr. S.K. Jha	Food Science & Post Harvest Technology	ICAR-IARI
52.	Dr. Shiv Dhar Mishra	Agronomy	ICAR-IARI
53.	Dr. D.K. Singh	Agricultural Engineering	ICAR-IARI
54.	Dr. S. Naresh Kumar	Environmental Sciences; **Nodal officer, Environmental Management Framework**	ICAR-IARI

Foreword

The Division of Agricultural Economics, a constituent of the School of Social Sciences of ICAR-Indian Agricultural Research Institute, was established in 1960. The mandate of the Division is to conduct research in frontier areas and serve as a centre for academic excellence in post-graduate education. Since its inception, the Division has been making contributions in basic and applied research with significant implications for agricultural policy. The Division has achieved excellence in post-graduate education and research as an ICAR-UNDP Centre of Excellence through a faculty exchange program for human resources development and strengthening of infrastructure facilities. Since 1995 it has been functioning as an ICAR Centre of Advanced Faculty Training (CAFT) to strengthen the capacity for agricultural economics and policy research in the national agricultural research system.

The research contributions of the Division have been globally recognized and many of the alumni occupy positions of repute in national and international organizations. The Division has maintained good academic liaison with other divisions at IARI and other national and international agricultural research institutions. The research focus of the Division has been continuously reoriented to address contemporary development challenges. Current research thrust areas of the division include investment in agriculture, inclusive growth, and poverty alleviation, the impact of agricultural technologies and policies, price forecasting and market outlooks, natural resource use in agriculture and ecosystem services, climate change effects, mitigation and adaptations, and food and nutritional security.

Use of the latest research methods and analytical tools in its research activities has always been the strength of the Division, which sets it apart. Considering the core strength of the Division in research methodology, a short online training course on "Advances in Research Methodology for Social Sciences" is being organized by the Division. The objective of this training program, sponsored by the Centre for Advanced Agricultural Science and Technology (CAAST) component of the World Bank-funded National Agricultural Higher Education Project (NAHEP), is to upgrade the research skills of the students of social science discipline. The training program aptly covers a range of topics including conducting computer-aided surveys, synthesising evidence from scientific literature, handling large data sets and software, neural networks and their applications, choice analysis, social network analysis, content analysis, research writing etc. I am sure that the lectures on various research methodologies and practical sessions will be of immense help to the participants.

Rashmi Aggarwal

Dean and Joint Director(Edn)
ICAR-IARI,New Delhi

Date: 30.08.2020

Preface

Qualitative and quantitative methods are essential components of evidence-based research in Social Sciences. Since the last two decades have experienced rapid advancement in the methodology and analytical techniques, as well as their applications in the field of social sciences, it becomes imperative to disseminate the knowledge of these novel techniques to the students. This training manual is prepared considering the target of upgrading the research skills of the post-graduate students of social sciences. The manual is based on the NAHEP-CAAST sponsored online short training course titled "Advances in Research Methodology for Social Sciences" organized by the Division of Agricultural Economics, ICAR-Indian Agricultural Research Institute, New Delhi. Centre for Advanced Agricultural Science and Technology (CAAST) is a new initiative and student-centric sub-component of World Bank sponsored National Agricultural Higher Education Project (NAHEP), granted to IARI to provide a platform for strengthening education and research activities of post-graduate students.

Social science research, particularly in the applied disciplines of Agricultural Economics, Agricultural Extension and Agricultural Statistics, is characterised by a diversity of theoretical perspectives, substantive orientation, methodological strategies, data collection practices and analytical techniques. The students of these disciplines usually have to face challenges in research, since it involves conceptualizing the problems relevant to the stakeholders, collecting and handling large data sets (both primary as well as secondary), choosing appropriate methodology (qualitative and quantitative), executing the analysis using appropriate statistical packages, and interpreting and presenting the results in a meaning and useful format to all: farmers, academia and policymakers.

Recognizing the duty to impart essential research skills to the social science students, we have taken up the task of conducting the training and preparing this manual on Advances in Research Methodology for Social Sciences. The various chapters of this manual are contributed by the eminent social scientists of the country, with expertise in analytical methods. In addition to the basic research methods, the manual also covers topics like conducting computer-aided surveys, synthesising evidence from scientific literature, handling large data sets and software, neural networks and their applications, choice analysis, social network analysis, content analysis, research writing etc. We take this opportunity to sincerely acknowledge the contribution of all the authors in the preparation of this manual. Considering the diversity and comprehensive nature of the topics covered, the manual can act as a quick and effective reference source for the students in their future research endeavours.

Alka Singh
R.N. Padaria
R.R. Burman
Aditya K.S.
Praveen K.V.

Date: 30.08.2020

Acknowledgments

1. Secretary DARE and Director General ICAR, New Delhi
2. Deputy Director General (Education), ICAR, New Delhi
3. Assistant Director General (HRD), ICAR, New Delhi
4. National Coordinator, NAHEP, ICAR, New Delhi
5. CAAST Team, ICAR-IARI, New Delhi
6. P.G. School, ICAR-IARI, New Delhi
7. Director, ICAR-IARI, New Delhi
8. Dean & Joint Director (Education), ICAR-IARI, New Delhi
9. Joint Director (Research), ICAR-IARI, New Delhi
10. Head, Division of Agriculture Economics, ICAR-IARI, New Delhi
11. Professor, Division of Agriculture Economics, ICAR-IARI, New Delhi
12. AKMU, ICAR-IARI, New Delhi
13. Staff & Students, Division of Agriculture Economics, ICAR-IARI, New Delhi

Contents

A Brief Prelude to Systematic Reviews and Meta-Analysis

Praveen K.V.
Division of Agriculture Economics, ICAR-Indian Agricultural Research Institute, New Delhi

Systematic reviews are a sort of literature review that utilizes systematic methods to gather secondary data and blend or synthesise the research evidence qualitatively or quantitatively. With the volume of research evidence on any topic growing at an ever-expanding rate, it is very difficult for individual researchers or policymakers to survey this tremendous amount of literature and arrive at the best decision on its basis. Following a systematic approach, systematic reviews help summarize the research knowledge on an intervention. It endeavours to gather all the empirical evidence that fits pre-determined eligibility criteria to answer to a particular research question. It utilizes systematic techniques that are chosen with the end goal of minimizing bias and hence giving more dependable findings from which conclusions can be drawn and choices made (Antman et al 1992, Oxman and Guyatt 1993).

Research questions

As in the case of any research, the first and most significant choice in setting up a systematic review is to decide its core interest. This is best done by framing the questions that the review looks to answer. Well-formulated questions will guide the systematic review procedure, including deciding eligibility criteria, literature search, gathering data from selected publications, organizing and presenting the findings (Cooper 1984, Hedges 1994, Oliver et al 2017). The FINER standards have been proposed to make life easy for a researcher while creating research questions. As per this strategy, questions ought to be Feasible, Interesting, Novel, Ethical, and Relevant (Cummings et al 2007). These measures raise key issues to be considered at the start of the review and ought to be borne as a primary concern when questions are framed.

A systematic review can address any research question that can be answered by primary research. Studies that compare interventions utilize the outcome of the participants to arrive at the impacts of various interventions. Statistical synthesis (for example meta-analysis) centres on comparison of a new intervention with the control. The differentiation between the outcomes of two groups treated contrastingly is known as the ‘effect’ or the 'treatment effect’. The primary objective of systematic reviews should be ideally framed in a single sentence. The objective can be structured as: 'To evaluate the impacts of [intervention or technology] for [income enhancement] in [types of individuals, region, and setting if specified]'. This may be trailed by at least one secondary targets, for instance identifying with various participant groups, varying comparison of interventions or diverse outcome measures. The detailing of review question(s) requires thought of a few key segments (Richardson et al 1995, Counsell 1997) which can be conceptualized by the 'PICO', an abbreviation for Population, Intervention, Comparison(s) and Outcome. The scope of the review should be just apt. It should not be too broad or narrow to be relevant.

Table 1. PICO formulation

Item	Example
Population	Farmers in developing countries Farmers involved in farmer groups or producer companies
Intervention	GM crops Integrated Pest Management
Comparator	Communities/famers not participating in FFS Farmers/communities receiving alternative interventions
Outcome	Yield Net revenue

Defining inclusion criteria

One of the highlights that differentiate a systematic review from a narrative review is that the authors of systematic review ought to pre-indicate criteria and standards for including and barring individual studies. When building up the protocol, one of the initial steps is to decide the components of the review question (the population, intervention(s), comparator(s) and outcome, or PICO components) and how the intervention, in the identified population, creates the outcomes. Eligibility criteria depend on the PICO components in addition to a specification of the kinds of studies that have addressed these inquiries. The population, intervention, and comparators in the review question can be usually translated into the inclusion criteria, but not always directly.

Literature search and study selection

Systematic reviews require a careful, objective, and reproducible search of a variety of sources to extract as many studies (eligible) as possible. The quest for studies should be as broad as possible to diminish the danger of reporting bias and to identify maximum evidence as possible. Database determination ought to be guided by the survey theme. 'Grey literature' should also be considered. Authors ought to search for dissertations and conference abstracts also. They should also think about looking through the web, hand searching of journals and looking through full texts of journals electronically where accessible. They ought to inspect past reviews on a similar theme and check reference lists of included studies. Suitable search strategy should be formulated for searching in different databases. Choices about which studies to include for a review is among the most compelling choices that are made in the review procedure and they include judgment. Involvements of at least two individuals, working independently, are required to decide if each study meets the qualification standards. A PRISMA flow chart mentioning the selection of studies at each stage should be included in the report.

Table 2. List of databases to search

Sl No.	Database
1	Web of Science (Social science citation index)
2	CeRA
3	Google scholar
4	AgEcon search

5	Econlit
6	CAB abstract
7	Medline, Pubmed
8	ERIC

Coding and Data collection

Authors are urged to create layouts of tables and figures that will show up in the review to encourage the design of data collection forms. The way to effective data collection is to build simple-to-use forms and gather adequate and unambiguous information that present the source in an organized and structured way. Effort ought to be made to collect information required for meta-analysis. Data ought to be gathered and documented in a structure that permits future access and data sharing. Coding should provide for adding data in the following components:

- Study identification
- Intervention discriptives
- Process and implementation
- Context
- Popultion characteristics
- Research methods
- Effect size data
- Outcomes
- Subgroups

Effect measures

The kinds of outcome data that authors are probably going to experience are dichotomous data, continuous data, ordinal data, count or rate data and time-to-event data. The nature of the collected data determines the effect measures of intervention. Effect measures are statistical constructs that compare outcome data between two intervention groups. It is mainly of two distributed into two categories: ratio measures and difference measures. Estimates of effect describe the size of the intervention effect in terms of how diverse the outcome data were between the groups. For ratio effect measures, 1 indicates no distinction between the groups, while for difference measures, 0 indicates no distinction between the groups. Larger and smaller values than these ‘null’ values may suggest either benefit or harm of an intervention. The true effects of interventions very difficult to arrive at, and can only be assessed by the available studies. Estimates should thus be presented with uncertainty measures like confidence interval or standard error (SE). Examples of effect measures of dichotomous outcome data: Risk ratio, Odds ratio, Risk difference. Examples of effect measures of continuous outcome data: Mean difference, Standardised mean difference, Ratio of means

Meta-analysis

Meta-analysis can be considered as a key step in a systematic review. Meta-analysis involves deciding on the possibility of combining the results of selected studies. This procedure results in an overall statistic with a confidence interval that summarizes the effect of an intervention compared with the counterfactual. Meta-analysis is useful since they improve precision by including more information that smaller individual studies lack. To carry out a meta-analysis, at first, a summary statistic is computed for individual studies, to present the effect of the intervention in a uniform measure. Next, the individual study's intervention effects are statistically combined using a weighted average of the intervention effects estimated in the individual studies. Undertake random-effects meta-analysis if the studies are not all estimating the same intervention effect, but estimate intervention effects that follow a distribution across studies. On the other hand, if each study is estimating the same quantity, then a fixed-effect meta-analysis can be used. A confidence interval is derived that represents the precision of the summarized estimate. Meta-analysis can be carried out using two models:

- ***Fixed effect model***
    - Under the fixed-effect model we assume that all studies in the meta-analysis share a common (true) effect size.
    - Put another way, all factors that could influence the effect size are the same in all the studies, and therefore the true effect size is the same in all the studies.
    - Since all studies share the same true effect, it follows that the observed effect size varies from one study to the next only because of the random error inherent in each study.
    - If each study had an infinite sample size the sampling error would be zero and the observed effect for each study would be the same as the true effect.
    - In practice, of course, the sample size in each study is not infinite, and so there is sampling error and the effect observed in the study is not the same as the true effect.
    - The observed effect for any study is given by the population mean plus the sampling error in that study.
- ***Random effects model***
    - There is no reason to assume that studies are identical in the sense that the true effect size is exactly the same in all the studies.
    - We might not have assessed these covariates in each study.
    - If each study had an infinite sample size the sampling error would be zero and the observed effect for each study would be the same as the true effect for that study.
    - The sample size in any study is not infinite and therefore the sampling error is not zero. The observed effect for that study will be less than or greater than the true effect because of sampling error.
    - The distance between the overall mean and
    - the observed effect in any given study consists of two distinct parts: true variation in effect sizes (i) and sampling error
    - The observed effect for any study is given by the grand mean, the deviation of the study's true effect from the grand mean, and the deviation of the study's observed effect from the study's true effect.

Meta-analysis: Demonstration (Example of meta-analysis of biofertilizer in India)

Setting the question

The effects of biofertilizer use in crop yields in India

- PICO
  - P- Experimental plots with biofertilizer application
  - I- Biofertilizer
  - C- Control plots
  - O- Yield

Search strategy for meta-analysis

A comprehensive literature search was undertaken from February to April 2019 in the google scholar, and CeRA (Consortium for e-resources in agriculture) to identify the studies to be included in the meta-analysis. The studies published between 2000 and 2019 were searched using the following search strings: "biofertilizer", "biofertiliser", "biofertilizer OR biofertiliser" AND "response" AND "India".

Screening, coding and data extraction

The studies were screened independently by authors to select the ones that meet the criteria to be included for the meta-analysis. The studies based on field trials, and that provide data for pairwise comparison of the yield effect of biofertilizer treated crop to that of the control are included. Full papers were reviewed to record the data on mean yields, standard deviations and the number of replications, and also other field-specific observations that would be required for the analysis. Out of the 16700 studies that appeared during the literature search, only 236 were selected after the preliminary screening to remove studies based on biofertilizer production technology, studies that are carried out in other countries, review studies, studies dealing with regulation and policies, and other aspects of biofertilizers that are not of our interest (these are termed as ‘exclusion criteria). After removing the duplicate studies and the ones based on laboratory experiments, 86 studies were selected for full-text reviews. From this, only 18 articles were finally selected for the meta-analysis, as the others did not provide the information that we require for meta-analysis.

The flow of the search process is given in detail in the Preferred Reporting Items for Systematic Reviews and Meta-Analyses(PRISMA) flow chart given in figure . The data from all the selected studies were then extracted and classified on the basis of types of biofertilizer. Nitrogen-fixing, phosphate solubilising, VAM, Combined biofertilizers, and others were the biofertilizer categories on the basis of which data extracted from the studies were grouped. Suitable predetermined codes were prepared in advance for this purpose. Example of coded sheet is given in the figure below. Further on the basis of crop groups, data were classified into that of cereals, legumes, vegetables and oilseeds. Thus from the 18 studies selected for meta-analysis, we were able to carry out 38 pairwise comparisons between biofertilizer treatment and control.

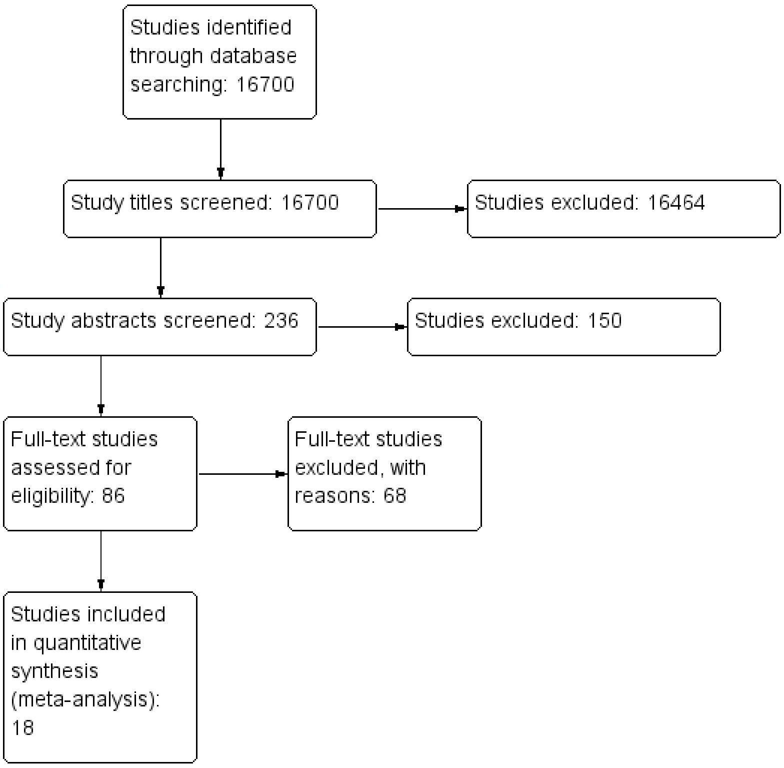
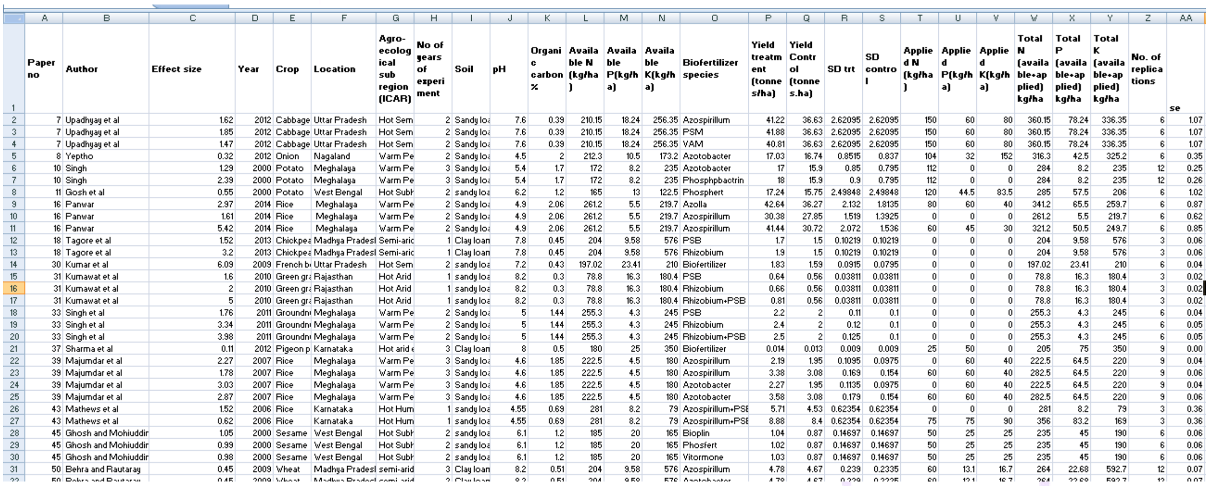

Studies identified through database searching: 16700
Study titles screened: 16700
Studies excluded: 16464
Study abstracts screened: 236
Studies excluded: 150
Full-text studies assessed for eligibility: 86
Full-text studies excluded, with reasons: 68
Studies included in quantitative synthesis (meta-analysis): 18

Figure 1. PRISMA flow chart

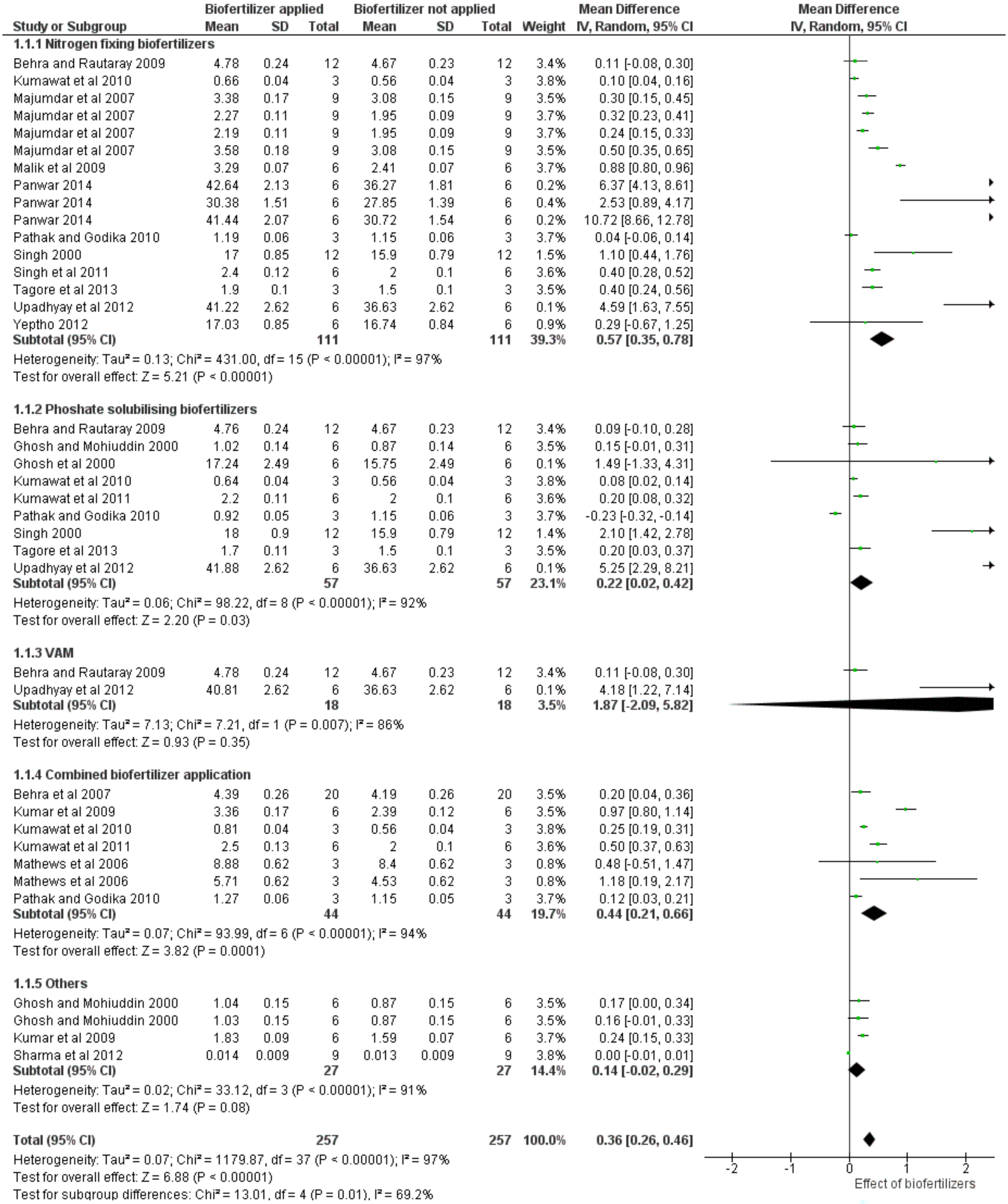


	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA
1	Paper no	Author	Effect size	Year	Crop	Location	Agro-ecological sub region (ICAR)	No of years of experiment	Soil	pH	Organic carbon %	Available N (kg/ha)	Available P(kg/ha)	Available K(kg/ha)	Biofertilizer species	Yield treatment (tonnes/ha)	Yield Control (tonnes.ha)	SD trt	SD control	Applied N (kg/ha)	Applied P(kg/ha)	Applied K(kg/ha)	Total N (available+applied) kg/ha	Total P (available+applied) kg/ha	Total K (available+applied) kg/ha	No. of replications	se
2	7	Upadhyay et al	1.62	2012	Cabbage	Uttar Pradesh	Hot Sem	2	Sandy loa	7.6	0.39	210.15	18.24	256.35	Azospirillum	41.22	36.63	2.62095	2.62095	150	60	80	360.15	78.24	336.35	6	1.07
3	7	Upadhyay et al	1.85	2012	Cabbage	Uttar Pradesh	Hot Sem	2	Sandy loa	7.6	0.39	210.15	18.24	256.35	PSM	41.88	36.63	2.62095	2.62095	150	60	80	360.15	78.24	336.35	6	1.07
4	7	Upadhyay et al	1.47	2012	Cabbage	Uttar Pradesh	Hot Sem	2	Sandy loa	7.6	0.39	210.15	18.24	256.35	VAM	40.81	36.63	2.62095	2.62095	150	60	80	360.15	78.24	336.35	6	1.07
5	8	Yeptho	0.32	2012	Onion	Nagaland	Warm Pe	2	Sandy loa	4.5	2	212.3	10.5	173.2	Azotobacter	17.03	16.74	0.8515	0.837	104	32	152	316.3	42.5	325.2	6	0.35
6	10	Singh	1.29	2000	Potato	Meghalaya	Warm Pe	3	Sandy loa	5.4	1.7	172	8.2	235	Azotobacter	17	15.9	0.85	0.795	112	0	0	284	8.2	235	12	0.25
7	10	Singh	2.39	2000	Potato	Meghalaya	Warm Pe	3	Sandy loa	5.4	1.7	172	8.2	235	Phosphpbactrin	18	15.9	0.9	0.795	112	0	0	284	8.2	235	12	0.26
8	11	Gosh et al	0.55	2000	Potato	West Bengal	Hot Subh	2	sandy loa	6.2	1.2	165	13	122.5	Phosphert	17.24	15.75	2.49848	2.49848	120	44.5	83.5	285	57.5	206	6	1.02
9	16	Panwar	2.97	2014	Rice	Meghalaya	Warm Pe	2	Sandy loa	4.9	2.06	261.2	5.5	219.7	Azolla	42.64	36.27	2.132	1.8135	80	60	40	341.2	65.5	259.7	6	0.87
10	16	Panwar	1.61	2014	Rice	Meghalaya	Warm Pe	2	Sandy loa	4.9	2.06	261.2	5.5	219.7	Azospirillum	30.38	27.85	1.519	1.3925	0	0	0	261.2	5.5	219.7	6	0.62
11	16	Panwar	5.42	2014	Rice	Meghalaya	Warm Pe	2	Sandy loa	4.9	2.06	261.2	5.5	219.7	Azospirillum	41.44	30.72	2.072	1.536	60	45	30	321.2	50.5	249.7	6	0.85
12	18	Tagore et al	1.52	2013	Chickpea	Madhya Pradesl	Semi-aric	1	Clay loan	7.8	0.45	204	9.58	576	PSB	1.7	1.5	0.10219	0.10219	0	0	0	204	9.58	576	3	0.06
13	18	Tagore et al	3.2	2013	Chickpea	Madhya Pradesl	Semi-aric	1	Clay loan	7.8	0.45	204	9.58	576	Rhizobium	1.9	1.5	0.10219	0.10219	0	0	0	204	9.58	576	3	0.06
14	30	Kumar et al	6.09	2009	French b	Uttar Pradesh	Hot Sem	2	sandy loa	7.2	0.43	197.02	23.41	210	Biofertilizer	1.83	1.59	0.0915	0.0795	0	0	0	197.02	23.41	210	6	0.04
15	31	Kumawat et al	1.6	2010	Green gra	Rajasthan	Hot Arid	1	sandy loa	8.2	0.3	78.8	16.3	180.4	PSB	0.64	0.56	0.03811	0.03811	0	0	0	78.8	16.3	180.4	3	0.02
16	31	Kumawat et al	2	2010	Green gra	Rajasthan	Hot Arid	1	sandy loa	8.2	0.3	78.8	16.3	180.4	Rhizobium	0.66	0.56	0.03811	0.03811	0	0	0	78.8	16.3	180.4	3	0.02
17	31	Kumawat et al	5	2010	Green gra	Rajasthan	Hot Arid	1	sandy loa	8.2	0.3	78.8	16.3	180.4	Rhizobium+PSB	0.81	0.56	0.03811	0.03811	0	0	0	78.8	16.3	180.4	3	0.02
18	33	Singh et al	1.76	2011	Groundn	Meghalaya	Warm Pe	2	Sandy loa	5	1.44	255.3	4.3	245	PSB	2.2	2	0.11	0.1	0	0	0	255.3	4.3	245	6	0.04
19	33	Singh et al	3.34	2011	Groundn	Meghalaya	Warm Pe	2	Sandy loa	5	1.44	255.3	4.3	245	Rhizobium	2.4	2	0.12	0.1	0	0	0	255.3	4.3	245	6	0.05
20	33	Singh et al	3.98	2011	Groundn	Meghalaya	Warm Pe	2	Sandy loa	5	1.44	255.3	4.3	245	Rhizobium+PSB	2.5	2	0.125	0.1	0	0	0	255.3	4.3	245	6	0.05
21	37	Sharma et al	0.11	2012	Pigeon p	Karnataka	Hot arid e	3	Clay loan	8	0.5	180	25	350	Biofertilizer	0.014	0.013	0.009	0.009	25	50	0	205	75	350	9	0.00
22	39	Majumdar et al	2.27	2007	Rice	Meghalaya	Warm Pe	3	Sandy loa	4.6	1.85	222.5	4.5	180	Azospirillum	2.19	1.95	0.1095	0.0975	0	60	40	222.5	64.5	220	9	0.04
23	39	Majumdar et al	1.78	2007	Rice	Meghalaya	Warm Pe	3	Sandy loa	4.6	1.85	222.5	4.5	180	Azospirillum	3.38	3.08	0.169	0.154	60	60	40	282.5	64.5	220	9	0.06
24	39	Majumdar et al	3.03	2007	Rice	Meghalaya	Warm Pe	3	Sandy loa	4.6	1.85	222.5	4.5	180	Azotobacter	2.27	1.95	0.1135	0.0975	0	60	40	222.5	64.5	220	9	0.04
25	39	Majumdar et al	2.87	2007	Rice	Meghalaya	Warm Pe	3	Sandy loa	4.6	1.85	222.5	4.5	180	Azotobacter	3.58	3.08	0.179	0.154	60	60	40	282.5	64.5	220	9	0.06
26	43	Mathews et al	1.52	2006	Rice	Karnataka	Hot Hum	1	sandy loa	4.55	0.69	281	8.2	79	Azospirillum+PSE	5.71	4.53	0.62354	0.62354	0	0	0	281	8.2	79	3	0.36
27	43	Mathews et al	0.62	2006	Rice	Karnataka	Hot Hum	1	sandy loa	4.55	0.69	281	8.2	79	Azospirillum+PSE	8.88	8.4	0.62354	0.62354	75	75	90	356	83.2	169	3	0.36
28	45	Ghosh and Mohiuddir	1.05	2000	Sesame	West Bengal	Hot Subh	2	sandy loa	6.1	1.2	185	20	165	Bioplin	1.04	0.87	0.14697	0.14697	50	25	25	235	45	190	6	0.06
29	45	Ghosh and Mohiuddir	0.99	2000	Sesame	West Bengal	Hot Subh	2	sandy loa	6.1	1.2	185	20	165	Phosfert	1.02	0.87	0.14697	0.14697	50	25	25	235	45	190	6	0.06
30	45	Ghosh and Mohiuddir	0.98	2000	Sesame	West Bengal	Hot Subh	2	sandy loa	6.1	1.2	185	20	165	Vitormone	1.03	0.87	0.14697	0.14697	50	25	25	235	45	190	6	0.06
31	50	Behra and Rautaray	0.45	2009	Wheat	Madhya Pradesl	semi-arid	3	Clay loan	8.2	0.51	204	9.58	576	Azospirillum	4.78	4.67	0.239	0.2335	60	13.1	16.7	264	22.68	592.7	12	0.07
32	50	Behra and Rautaray	0.45	2009	Wheat	Madhya Pradesl	semi-arid	3	Clay loan	8.2	0.51	204	9.58	576	Azotobacter	4.78	4.67	0.239	0.2335	60	13.1	16.7	264	22.68	592.7	12	0.07

Figure 2. Coding

Meta-analysis

Mean difference was selected as the effect size. As per the results of the meta-analysis, application of biofertilizers resulted in an average yield increase of 0.36 tonnes per ha in India. The diamond shape gives the effect of subgroup and total biofertilizers. The size of the diamond shape gives the magnitude of the effect size and the edges represent the confidence interval (95% level). Meta-regerssion results suggest that only the combined biofertilizer application has a significant effect on yield improvement. The model, indicated significant yield increase due to biofertilizers in clay loam soil (in comparison to sandy loam), and soils with low K and high P content as well as low pH and low organic carbon content (in line with the findings of Schults, 2018). The variation in the performance of biofertilizers as per the agro-ecological conditions was also confirmed in this model. Most agro-ecological variables considered were significant. Among the crop groups, significant yield effects were detected in the case of cereals, legumes and vegetables (first model).



	Biofertilizer applied			Biofertilizer not applied				Mean Difference
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	Weight	IV, Random, 95% CI
1.1.1 Nitrogen fixing biofertilizers								
Behra and Rautaray 2009	4.78	0.24	12	4.67	0.23	12	3.4%	0.11 [-0.08, 0.30]
Kumawat et al 2010	0.66	0.04	3	0.56	0.04	3	3.8%	0.10 [0.04, 0.16]
Majumdar et al 2007	3.38	0.17	9	3.08	0.15	9	3.5%	0.30 [0.15, 0.45]
Majumdar et al 2007	2.27	0.11	9	1.95	0.09	9	3.7%	0.32 [0.23, 0.41]
Majumdar et al 2007	2.19	0.11	9	1.95	0.09	9	3.7%	0.24 [0.15, 0.33]
Majumdar et al 2007	3.58	0.18	9	3.08	0.15	9	3.5%	0.50 [0.35, 0.65]
Malik et al 2009	3.29	0.07	6	2.41	0.07	6	3.7%	0.88 [0.80, 0.96]
Panwar 2014	42.64	2.13	6	36.27	1.81	6	0.2%	6.37 [4.13, 8.61]
Panwar 2014	30.38	1.51	6	27.85	1.39	6	0.4%	2.53 [0.89, 4.17]
Panwar 2014	41.44	2.07	6	30.72	1.54	6	0.2%	10.72 [8.66, 12.78]
Pathak and Godika 2010	1.19	0.06	3	1.15	0.06	3	3.7%	0.04 [-0.06, 0.14]
Singh 2000	17	0.85	12	15.9	0.79	12	1.5%	1.10 [0.44, 1.76]
Singh et al 2011	2.4	0.12	6	2	0.1	6	3.6%	0.40 [0.28, 0.52]
Tagore et al 2013	1.9	0.1	3	1.5	0.1	3	3.5%	0.40 [0.24, 0.56]
Upadhyay et al 2012	41.22	2.62	6	36.63	2.62	6	0.1%	4.59 [1.63, 7.55]
Yeptho 2012	17.03	0.85	6	16.74	0.84	6	0.9%	0.29 [-0.67, 1.25]
Subtotal (95% CI)			**111**			**111**	**39.3%**	**0.57 [0.35, 0.78]**
Heterogeneity: Tau² = 0.13; Chi² = 431.00, df = 15 (P < 0.00001); I² = 97%								
Test for overall effect: Z = 5.21 (P < 0.00001)								
1.1.2 Phoshate solubilising biofertilizers								
Behra and Rautaray 2009	4.76	0.24	12	4.67	0.23	12	3.4%	0.09 [-0.10, 0.28]
Ghosh and Mohiuddin 2000	1.02	0.14	6	0.87	0.14	6	3.5%	0.15 [-0.01, 0.31]
Ghosh et al 2000	17.24	2.49	6	15.75	2.49	6	0.1%	1.49 [-1.33, 4.31]
Kumawat et al 2010	0.64	0.04	3	0.56	0.04	3	3.8%	0.08 [0.02, 0.14]
Kumawat et al 2011	2.2	0.11	6	2	0.1	6	3.6%	0.20 [0.08, 0.32]
Pathak and Godika 2010	0.92	0.05	3	1.15	0.06	3	3.7%	-0.23 [-0.32, -0.14]
Singh 2000	18	0.9	12	15.9	0.79	12	1.4%	2.10 [1.42, 2.78]
Tagore et al 2013	1.7	0.11	3	1.5	0.1	3	3.5%	0.20 [0.03, 0.37]
Upadhyay et al 2012	41.88	2.62	6	36.63	2.62	6	0.1%	5.25 [2.29, 8.21]
Subtotal (95% CI)			**57**			**57**	**23.1%**	**0.22 [0.02, 0.42]**
Heterogeneity: Tau² = 0.06; Chi² = 98.22, df = 8 (P < 0.00001); I² = 92%								
Test for overall effect: Z = 2.20 (P = 0.03)								
1.1.3 VAM								
Behra and Rautaray 2009	4.78	0.24	12	4.67	0.23	12	3.4%	0.11 [-0.08, 0.30]
Upadhyay et al 2012	40.81	2.62	6	36.63	2.62	6	0.1%	4.18 [1.22, 7.14]
Subtotal (95% CI)			**18**			**18**	**3.5%**	**1.87 [-2.09, 5.82]**
Heterogeneity: Tau² = 7.13; Chi² = 7.21, df = 1 (P = 0.007); I² = 86%								
Test for overall effect: Z = 0.93 (P = 0.35)								
1.1.4 Combined biofertilizer application								
Behra et al 2007	4.39	0.26	20	4.19	0.26	20	3.5%	0.20 [0.04, 0.36]
Kumar et al 2009	3.36	0.17	6	2.39	0.12	6	3.5%	0.97 [0.80, 1.14]
Kumawat et al 2010	0.81	0.04	3	0.56	0.04	3	3.8%	0.25 [0.19, 0.31]
Kumawat et al 2011	2.5	0.13	6	2	0.1	6	3.6%	0.50 [0.37, 0.63]
Mathews et al 2006	8.88	0.62	3	8.4	0.62	3	0.8%	0.48 [-0.51, 1.47]
Mathews et al 2006	5.71	0.62	3	4.53	0.62	3	0.8%	1.18 [0.19, 2.17]
Pathak and Godika 2010	1.27	0.06	3	1.15	0.05	3	3.7%	0.12 [0.03, 0.21]
Subtotal (95% CI)			**44**			**44**	**19.7%**	**0.44 [0.21, 0.66]**
Heterogeneity: Tau² = 0.07; Chi² = 93.99, df = 6 (P < 0.00001); I² = 94%								
Test for overall effect: Z = 3.82 (P = 0.0001)								
1.1.5 Others								
Ghosh and Mohiuddin 2000	1.04	0.15	6	0.87	0.15	6	3.5%	0.17 [0.00, 0.34]
Ghosh and Mohiuddin 2000	1.03	0.15	6	0.87	0.15	6	3.5%	0.16 [-0.01, 0.33]
Kumar et al 2009	1.83	0.09	6	1.59	0.07	6	3.7%	0.24 [0.15, 0.33]
Sharma et al 2012	0.014	0.009	9	0.013	0.009	9	3.8%	0.00 [-0.01, 0.01]
Subtotal (95% CI)			**27**			**27**	**14.4%**	**0.14 [-0.02, 0.29]**
Heterogeneity: Tau² = 0.02; Chi² = 33.12, df = 3 (P < 0.00001); I² = 91%								
Test for overall effect: Z = 1.74 (P = 0.08)								
Total (95% CI)			**257**			**257**	**100.0%**	**0.36 [0.26, 0.46]**
Heterogeneity: Tau² = 0.07; Chi² = 1179.87, df = 37 (P < 0.00001); I² = 97%								
Test for overall effect: Z = 6.88 (P < 0.00001)								
Test for subgroup differences: Chi² = 13.01, df = 4 (P = 0.01), I² = 69.2%								

Figure 3. Forest plot

Table3. Meta-regression

Variables	Model with biofertilizer and agro-ecological groups	
	Coefficient	SE
Experiment duration	-6.99***	1.74
Ph	-2.50**	1.05
organic carbon	-2.03	1.76
Total N	0.00	0.01
Total P	0.03*	0.02
Total K	-0.04***	0.01
Replication number	1.35***	0.36
Clay loam	33.72***	8.42
VAM	-0.46	1.87
Combined biofertilizers	3.02**	1.25
Nitrogen fixers	0.51	1.11
Phosphate solubilizers	-0.38	1.01
Hot arid eco region	16.17**	5.81
Hot semi arid eco region	21.56***	4.65
Hot sub humid eco region	14.01***	3.50
Hot arid eco sub region	-8.67**	3.25
Northern plain	28.35***	5.00
Semi arid tropics	-1.18	1.51
Warm perhumid eco region	16.08***	3.66
Hot arid eco sub region	33.29***	6.84
Constant	16.54***	4.57
Observations	38	
R-squared adjusted	83.25	
F statistic	9.5	
Tau-sq	0.890	
I-sq	99.70	

References

Antman E, Lau J, Kupelnick B, Mosteller F, Chalmers T. A comparison of results of meta-analyses of randomized control trials and recommendations of clinical experts: treatment for myocardial infarction. JAMA 1992; 268: 240–248.

Cooper H. The problem formulation stage. In: Cooper H, editor. Integrating Research: A Guide for Literature Reviews. Newbury Park (CA) USA: Sage Publications; 1984.

Counsell C. Formulating questions and locating primary studies for inclusion in systematic reviews. Annals of Internal Medicine 1997; 127: 380–387.

Cummings SR, Browner WS, Hulley SB. Conceiving the research question and developing the study plan. In: Hulley SB, Cummings SR, Browner WS, editors. Designing Clinical Research: An Epidemiological Approach. 4th ed. Philadelphia (PA): Lippincott Williams & Wilkins; 2007. p. 14–22.

Hedges LV. Statistical considerations. In: Cooper H, Hedges LV, editors. The Handbook of Research Synthesis. New York (NY): USA: Russell Sage Foundation; 1994.

Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.0 (updated July 2019). Cochrane, 2019. Available from www.training.cochrane.org/handbook.

Oliver S, Dickson K, Bangpan M, Newman M. Getting started with a review. In: Gough D, Oliver S, Thomas J, editors. An Introduction to Systematic Reviews. London (UK): Sage Publications Ltd.; 2017.

Oxman A, Guyatt G. The science of reviewing research. Annals of the New York Academy of Sciences 1993; 703: 125–133.

Praveen KV, Singh A. Realizing the potential of a low-cost technology to enhance crop yields: evidence from a meta-analysis of biofertilizers in India. Agricultural Economics Research Review. 2019;32(conf):77-91.

Richardson WS, Wilson MC, Nishikawa J, Hayward RS. The well-built clinical question: a key to evidence-based decisions. ACP Journal Club 1995; 123: A12 –13

Disclaimer: This chapter of the manual is prepared referring the sources listed in the reference section. This is prepared only for distribution to the participants of the training. Kindly cite the original sources while quoting the contents of this chapter.

Lecture Notes on Basic Stata Commands

Aditya K. S.
Division of Agriculture Economics, ICAR-Indian Agricultural Research Institute, New Delhi

This chapter is intended to introduce stata to beginners. The commands indicated here are very basic commands keeping in mind a person who never used stata before. Advanced users can straight away attempt to attend the exercise section given at the end. Those who don't have access to Stata software can request a short term student licence by filling out this form https://www.stata.com/customer-service/short-term-license/. To begin with, stata allow us to directly type the command to perform any required analysis. Alternatively, there is also an interactive UI; drop-down menu. For example, to summarize a data, by typing summarize variable list, we can get the result. One can also do it by using the drop-down menu by going to current statistics. Nevertheless, it is better to use the commands rather than menu-based UI, as different commands can be saved in a one file called "do file" which enables single click execution. Few basic stata commands and functions are discussed below in detail.

The help command: Help command is probably the most useful command that every Stata user should know. It makes easier to get help on various commands. For example, one want to know about summarize command. In the command window, type

help summarize

All the details regarding the summarize command is listed, along with the examples. However, there is every possibility that I may not know the command name to use help command. In such case, I can use the search command. For example, one wants to run factor analysis and has no clue about the command to use for this analysis.

help factoranalysis

There is every possibility that stata returns with error message. Instead,

search factor analysis

Find it command: stata has many user written commands, which are made available as packages. These are verified by stata, and published in stata journal. These packages can simplify the complex analysis and enable us to use the methodology with single click. To find and install such user written packages, one can use findit command.

Findit psmatch2

Result sheet will include the stata journal articles and packages. Find the suitable one from the package list and install.

Simple mathematical operators

Mathematical operators are very similar to those used in MS excel

display 2+3

display (2/3)^2

Rational and logical operators

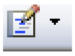
Rational operators include- >, <, >=,<=

Logical operators are ‘|’ *for ‘or’, ‘&’ for ‘and’*

Use of do files and log files

a) Dofiles in stata

Do files and log files are extremely useful for researchers. Do file is a stata file, which is used to store the commands. Usually, analysis involves many steps, with each step requiring use of certain set of commands. Storing them in form of do files will help to reduce the time. Do files will also help us to revisit the steps followed and treatment of variables. Do files are also used to share the commands with others.

Open the dofile from the stata drop down menu 

Open new dofile

Copy all the used commands and save do file

Commands in the do file are directly executable

b) Log files in stata

These days, many International journals demand the log files when we submit article for publications. Log files essentially record all our activities in a particular session. It includes the record of commands used and results obtained. It is a good practice to open a log file before starting an analysis.

Log using filename, format

Example: Log using demonstration.smcl

Browsing the data files: We can use the browse command to see the data set we are using. To refine our results browse can be combined with ‘if’ or ‘or’ or ‘and’ command.

Example : br if sex==1 & age==18 | age==17

String and numeric variables: Commonly, variables are saved in stata in either string or numeric format. Usually, the qualitative variables are saved in string format such as name, village name etc. Many commands not compatible with string variables. In such cases, destring command can be used. One should note that the values of string variables are enclosed in “”.

Example: destring Household_id, generate (hhid_new)

Handling the data set

1. Creating a new variable : Usually, gen or egen command is used to create the new variable. These commands can be combined with arithmetic operators or logical operators

Example : gen illiterate=1 if edu==1

egen gross_return_rank=rank(gross_return)

egen stdev_age= std(age)

2. The replace command: Replace command generally helps in editing value of already existing or generated variable, this command can be used.

   Example: replace sex_dummy=0 if missing(sex_dummy)

3. Sort command- to sort the data in ascending order. To order the data in descending order use gsort command

4. Tabulate command is used to make tables from the data.

5. Tabstat command: To tabulate the descriptive of the variables.

6. Keep or drop command can be used to delete the unwanted variable/ dropping few observations.

7. Tostring and destring : Command that can convert string to numeric variables

8. Xi command: Automatically creates dummy variables for the specified categorical variable.

Few Commands for Basic Statstical analysis

1. Regression : reg or probit or logit
2. Marginal Effects after probit or logit: mfx
3. Correlation: corr or pwcorr
4. Student T test: ttest
5. Chisquare test: tab, all
6. Principal Component Analysis: PCA
7. Factor analysis: factor

Few useful userwritten commands

1. tatable2: Calculate group wise mean value and test the significance
2. orth_out: Perform t-test for any number of variables at once
3. pscore: Estimates Propensity Scores
4. psmatch2: Perform Propensity Score Matching
5. cem: Perform Coarsened Exact Matching
6. doubleb: Perform Double Bound Contingent Valuation.
7. clustersampsi: perform power calculations for RCT

Exercise- 1

(write a ‘Do file’ for the following questions and submit @ adityaag68@gmail.com**.** You can also ping in case you have any questions/ clarifications**)**

1. Import the dataset “prod_data_12”, which has been shared with you already.
2. Summarize the data so that you see the means, standard deviation, min, and max of each variable
3. Find the correlation between Area and Yield, then create a simple plot of this relationship
4. Create a kernel density (kdensity) plot of yield per acre
5. Generate a new variable that indicates (i.e. 1 or 0) yields above 3,000 kg per acre.
6. Create a simple histogram of fertilizer expenditure
7. Create a more complex histogram of fertilizer expenditure by sex (i.e. overlapping histograms, one for males the other for females).
8. Create the following new variables:

- Total expenditure on fertilizer and pesticide
- The log of yield
- Numeric variable for sex called “male”
- Variable called “farmer” that indicates that the respondent’s primary occupation is farming
- Variable called “ownland” to indicate that the respondent owns the plots they farmed

9. Create dummy variables for a.) each soil type, and b.) each land type
10. Create a single variable that contains the average fertilizer expenditure in each village
11. Test that the average log yield between men and women is statistically different from one another
12. Regress log yield onto whatever variables you feel are most relevant

Exercise-II : Working with NSSO data using stata
Attempt this only after attending the session on handling the session on handling NSSO data

1. Understand the structure of AH0533V1 data set. Identify the merging strategy that can be adopted. (What should be i and j variables??)
2. Reshape the file AH0533V1.
3. Compute variables of interest (Total land, Total Quantity etc)
4. Merge the reshaped file with AH0233V1 (Which is the household data)
5. Characterize the wage earners in the dataset (Regress dummy for wage earning against set of explanatory variables (Include total value from crop as a variable- from the AH0533V2))
6. Use state fix effects and cluster the standard errors by NSS region to increase the precision of estimates

Computer Aided Personal Interviews

Aditya K. S.
Division of Agriculture Economics, ICAR-Indian Agricultural Research Institute, New Delhi

In this chapter we discuss survey data collection using computers, tablets and smart phones. Tablet/smart phones based surveys are better than paper based surveys as they make it easy to monitor the data collection on real time basis. Also, it saves time in data entry, as the collected data is directly retrieved to an Excel file. Here we briefly discuss approaches for collecting data with tablet/smartphones. Specifically, we provide detailed step by step explanation using Kobo toolbox, which is an open access resource for collecting survey data using tablets/mobile phones.

Context/Background:

Surveys are the bread and butter for any researchers in social sciences; be it for students (Masters and Ph.D) or researchers or organizations such as National Sample Survey Office (NSSO) which carryout large surveys for tracking different socio-economic characteristics. Research in the field of social science is mostly observational (except experimental approaches) and researcher has no control over the events. The data is collected as they are observed and analyzed to draw conclusions and the researcher has to depend on either secondary or primary data to draw inferences. As most of the secondary data sources are either macro level aggregates or not suitable to address the specific research question under considerations, primary surveys are inevitable in most research projects. Primary surveys offer the flexibility to the researcher to understand different socio-economic and cultural perspectives related to the research problem.

Paper based survey

Traditionally, the surveys use paper based questionnaire. Set of questions (Structured or semi-structured) are developed based on research problem and printed in papers (a lot of papers). The enumerators (surveyor) goes to the field and record the responses in the paper, which later gets digitized through data entry process. This is a cumbersome process as it takes quite a lot of time and has several problems and challenges associated with it. For instance, *data entry can take long time* and many *human errors can creep in*. Also, *real time monitoring is difficult* and even often the errors in surveys are detected once the data is entered and tabulated.

Computer-Aided Personal Interviews (CAPI)

Computer-Aided Personal Interviews (CAPI) is one such a technological advancement in data collection (Satellite and big data are the other advancements). CAPI is a face-to-face data collecting method in which the enumerator uses a small computer, or a tablet or a smart phone to collect the data. Availability of hardware (like cheap computers, tablets and smart phones) and software's (both paid and Open access) has made it easier to carry out CAPI surveys. This is helpful for two very important reasons; 1. No need for data entry process to digitize the information as in paper-based surveys. 2. Real time data monitoring to minimize the errors. There are many software tools available which are customized to carry out face-to-face surveys. There are wide variety of software tools to choose from, and they can be used by people with no programming knowledge (Like us!).

There are several advantages and disadvantages with CAPI which we discuss here (The advantages and disadvantaged are developed based on DIME 2020 and our own experience).

Advantages

- The collected data gets digitized immediately and sent to the server. In places with good networks we could carry out high frequency checks (cross checking the responses on a daily basis)- Real time monitoring is really easy unlike paper surveys, a person sitting at office can cross tabulate all the responses and test for data consistency.
- Could monitor the enumerators in case of larger surveys. The start time, end time, GPS location can be automatically recorded and supervisors/researchers go see it. Even the interviews could be randomly recorded (without knowledge of enumerators) and later verified by supervisors/researchers.
- It's easy in case of questions with skip logic; direct to a different set of question based on the response to a previous question (eg: If yes then…? And if No then… ?)
- Do calculations and conversations easily use inbuilt calculators. (For instance Biga to Hectare)
- Collect data such as images, farm plot size (GIS based plots), qualitative data (record statements)
- Avoid errors in data collection (missing questions) and data entry (most common errors)
- Data validation can be incorporated. For example, farming experience cannot be more than the age of the farmer. Such conditions can be imposed while preparing the survey itself to avoid the errors.

In nutshell, CAPI offers us many advantages in terms of features to reduce the data collection errors. Irrespective these advantages, there are certain dis-advantages rather challenges in using CAPI.

Disadvantages

- Respondents may not be comfortable with CAPI based survey. They get suspicious and often distracted seeing the gadgets.
- Need trained enumerators (Longer training periods) with a bit of knowledge of handling gadgets
- Difficult to carry computers and tablets in crime-ridden areas (Risk of theft- in one of the surveys it happened with us!)
- They also need electricity (Back up batteries could help) and good network connectivity
- Developing the questionnaire in CAPI is time consuming compared to paper based questionnaire (Though the overall time saved is positive)
- Language restrictions (Unlike paper based survey, CAPI has limited options for developing the questionnaire in local language).

Software's and Hardwares needs to carry out CAPI survey

We have different software's available for CAPI surveys. A list of commonly used ones are given in table 1. Each of them have their own set of advantages and disadvantages. A detailed discussion on each of them is beyond the scope of this blog. Keeping in mind that many of our readers would like to know about the open source options to carry out CAPI surveys, we have provided the detailed step by step guide to use Kobo toolbox; a free and open access software to carry out CAPI survey. However, if the researcher has money to go for paid softwares, he can always buy one.

Before starting our discussion on Kobo, let us try to understand few points to keep in mind while choosing the software and hardware for CAPI. Few important factors for selection of software are., What kind of data is required (text, pictures audio), how they are managed (requires own server?), output file format (most of them have multiple option), does it have language support (native language questionnaire) etc., Similarly, While choosing the hardware's(tablet/computer/smartphone), first thing is that the hardware should be compatible with the software requirements, it also need good quality camera (if pictures are to be recorded). Further the size of the screen depends on the questionnaire length (for smaller questionnaire smart phones are fine, but for larger questionnaires tablets with 7 inch screen is preferred). The hardware should also have enough internal memory (8 GB preferably) and external storage (SD cards), the gadget should also be GPS enabled with better accuracy (10-15 meters) and good battery life. It is a good practice to purchase tablets at the institute level. However, there is also option to rent them.

Table 1. Software's for CAPI

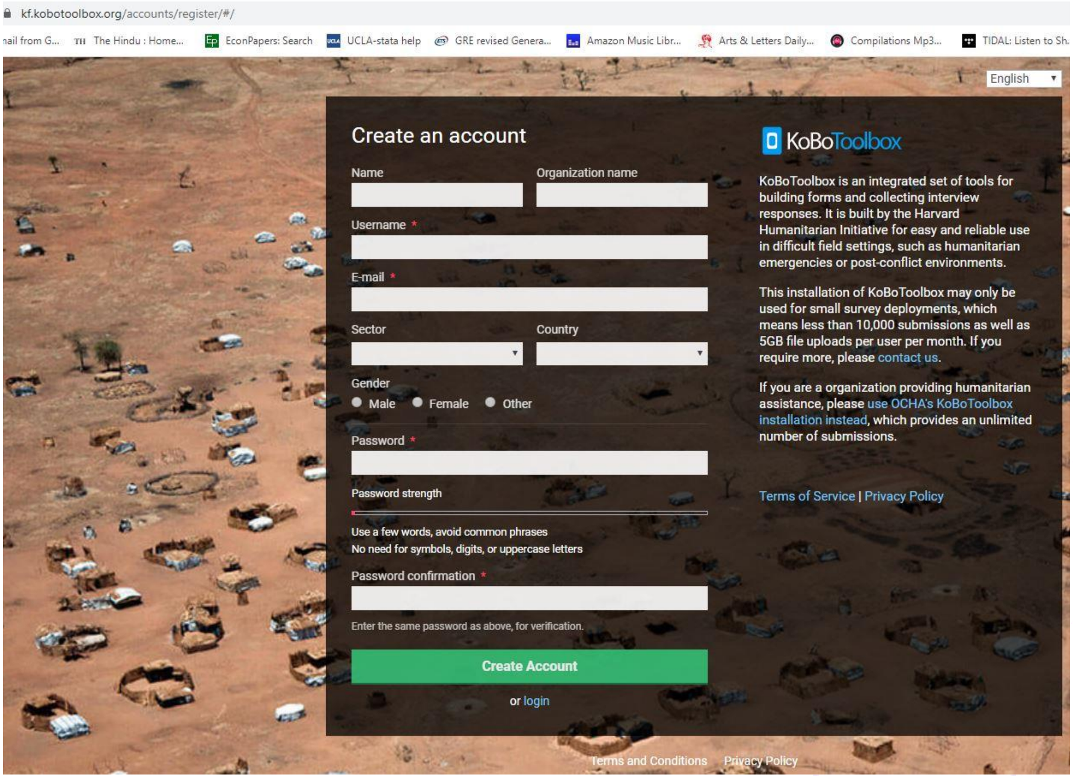
S.No.	Software	Developer	Access option	Link
1	Blaise	Statistics Netherlands	Restricted	https://blaise.com/
2	SurveyCTO	Dobility Inc	Paid	https://www.surveycto.com/
3	CSPro	United States Census Bureau	Open Access	https://www.census.gov/data/software/cspro.html
4	Dooblo	Dooblo Ltd., Israel	Paid	https://www.dooblo.net/
5	SurveyBe	Economic Development Initiatives Limited, UK	Paid	https://surveybe.com/
6	SurveySolutions	The World Bank	Open Access	https://mysurvey.solutions/
7	Kobo Tool Box	Harvard Humanitarian Initiative	Open access (Researchers)	https://www.kobotoolbox.org/
8	Open Data Kit (ODK)	University of Washington's Department of Computer Science and Engineering	Open access	https://docs.opendatakit.org/odk-x/survey-intro/

As stressed earlier, though many purchased software options are available, in this blog we will focus on one very commonly used, free and open access software support for CAPI survey- Kobo tool box.

Kobo Tool Box

Kobo toolbox is an open access toolset for collecting survey data using mobiles or tablets. In this following section we will elaborate on how to use kobo toolbox with a detailed step by step guide. Relevant screenshots are also provided.

- Step 1:Register as a researcher at kobo in the following link https://www.kobotoolbox.org/. This offers 10000 submissions per month with 5 GB of storage space, which is sufficient to most of the surveys done in academia. After registering, please note the username carefully, you will need it later.



Create an account

Name

Organization name

Username *

E-mail *

Sector

Country

Gender

○ Male ○ Female ○ Other

Password *

Password strength

Use a few words, avoid common phrases
No need for symbols, digits, or uppercase letters

Password confirmation *

Enter the same password as above, for verification.

Create Account

or login

KoBoToolbox is an integrated set of tools for building forms and collecting interview responses. It is built by the Harvard Humanitarian Initiative for easy and reliable use in difficult field settings, such as humanitarian emergencies or post-conflict environments.

This installation of KoBoToolbox may only be used for small survey deployments, which means less than 10,000 submissions as well as 5GB file uploads per user per month. If you require more, please contact us.

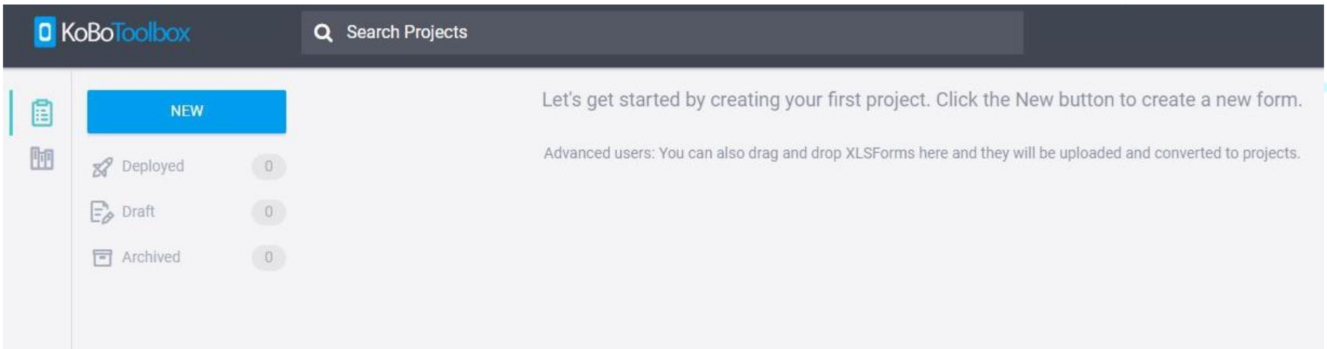
If you are a organization providing humanitarian assistance, please use OCHA's KoBoToolbox installation instead, which provides an unlimited number of submissions.

Terms of Service | Privacy Policy

Terms and Conditions Privacy Policy

Step 1

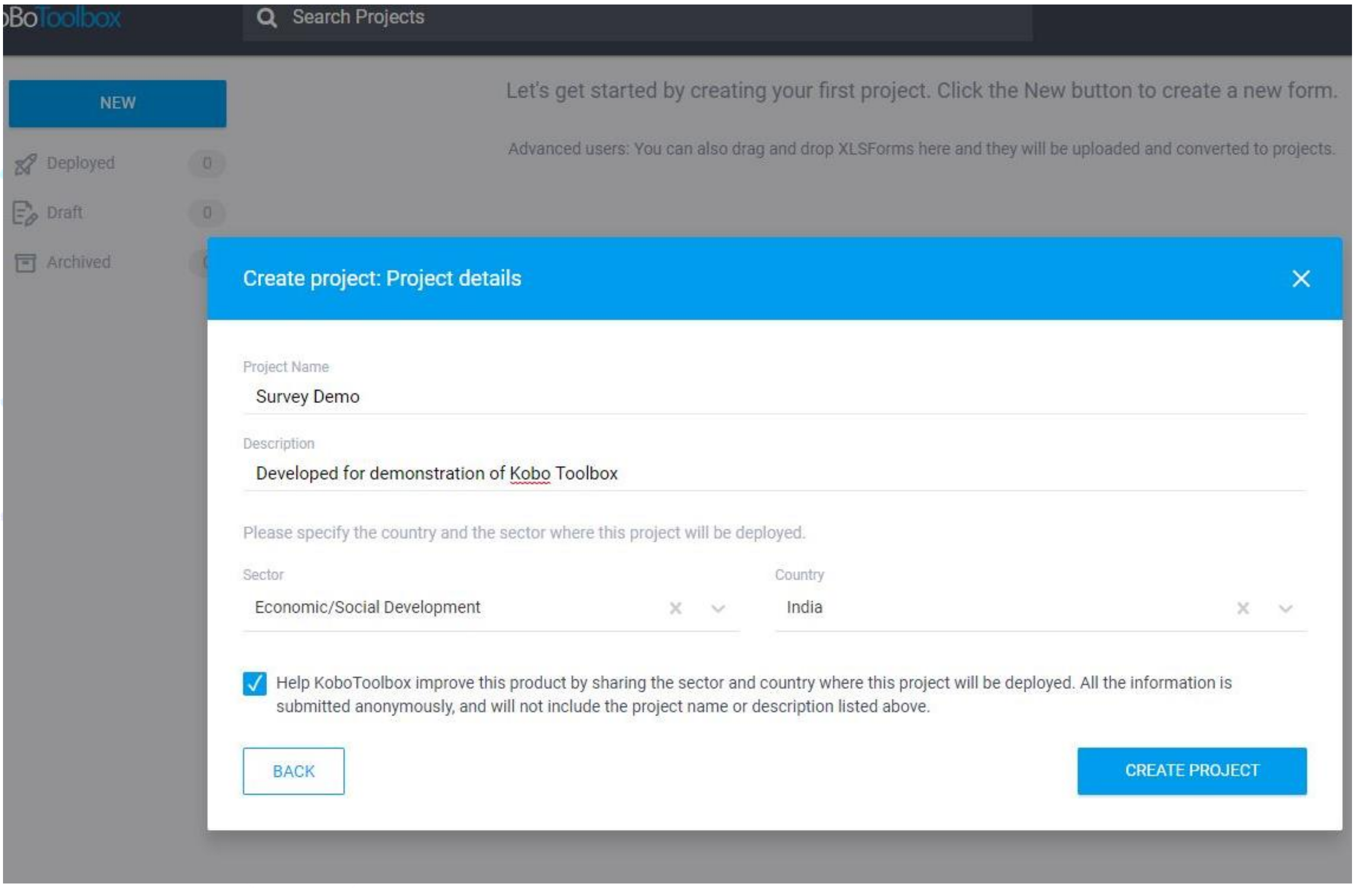
- Step 2: Login to your account and click on New to start preparing the survey schedule. Give a project name, select the suitable discipline and select the country. Click ‘Create Project’ to proceed to next step.



Let's get started by creating your first project. Click the New button to create a new form.

Advanced users: You can also drag and drop XLSForms here and they will be uploaded and converted to projects.

Step 2



Let's get started by creating your first project. Click the New button to create a new form.

Advanced users: You can also drag and drop XLSForms here and they will be uploaded and converted to projects.

Create project: Project details

Project Name: Survey Demo

Description: Developed for demonstration of Kobo Toolbox

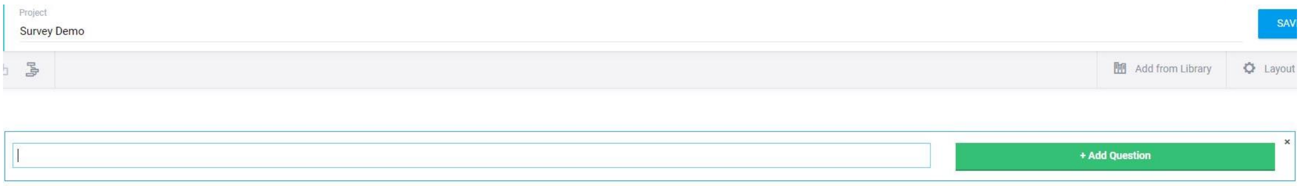
Please specify the country and the sector where this project will be deployed.

Sector: Economic/Social Development

Country: India

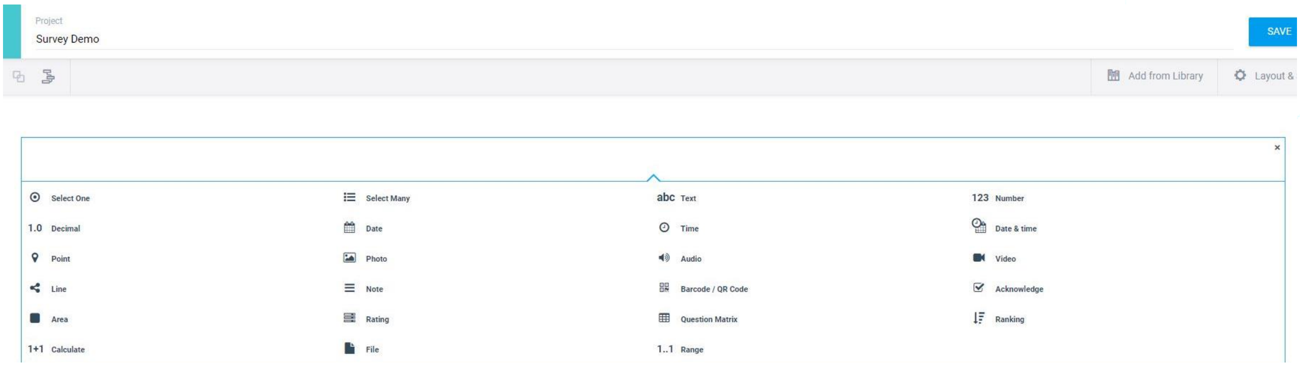
☑ Help KoboToolbox improve this product by sharing the sector and country where this project will be deployed. All the information is submitted anonymously, and will not include the project name or description listed above.

- Step 3: Click on +Add Question and select the question type. There are different type of questions available- text, numerical, select one, select many, decimal, rating, ranking, grid, date and time etc. According to the expected type of answer, select the right question type.



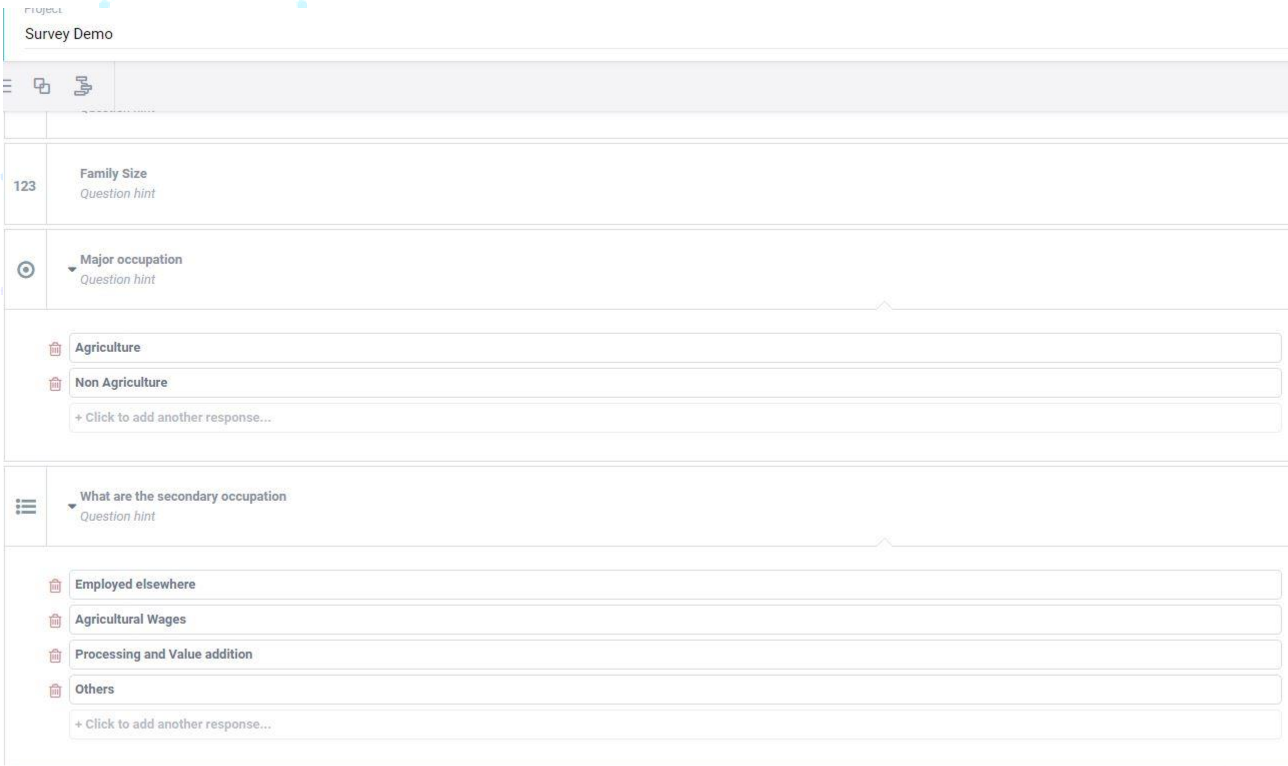
Project: Survey Demo

Step 3



Project: Survey Demo

- For example, family size will be expressed in terms of whole number, hence the question type has to be numeric. In the question ‘Major Occupation’, the right question type is select one, where either ‘Agriculture’ or “Non Agriculture’ has to be selected. In the ‘Secondary Occupation’ question, farmer can have more than one secondary occupations, hence, select many is the right choice.



Project

Survey Demo

123 Family Size

Question hint

Major occupation

Question hint

Agriculture

Non Agriculture

+ Click to add another response...

What are the secondary occupation

Question hint

Employed elsewhere

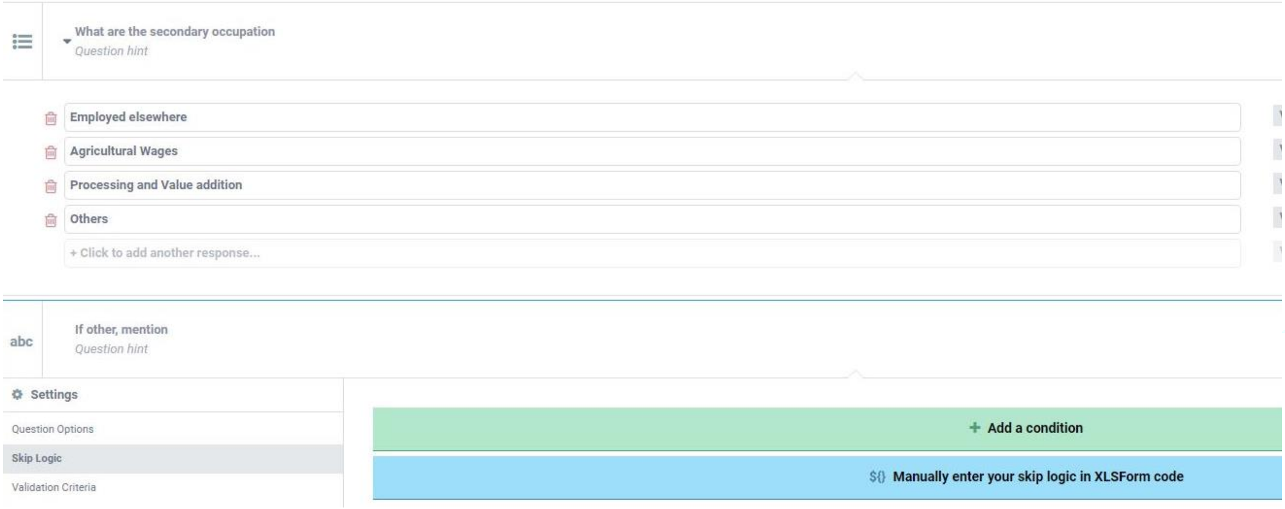
Agricultural Wages

Processing and Value addition

Others

+ Click to add another response...

- Skip logic is another attractive feature. Based on the response to previous question, you can impose a condition to display or skip a particular question. For instance, the question, mention the other occupation is displayed only if the respondent has selected the other occupation in the previous question. Or else that question will be skipped. This can save lot of time. For instance, think of survey involving two crops. If the farmer is not growing wheat, all the questions relating to wheat will be skipped if this logic function is appropriately used.



What are the secondary occupation

Question hint

Employed elsewhere

Agricultural Wages

Processing and Value addition

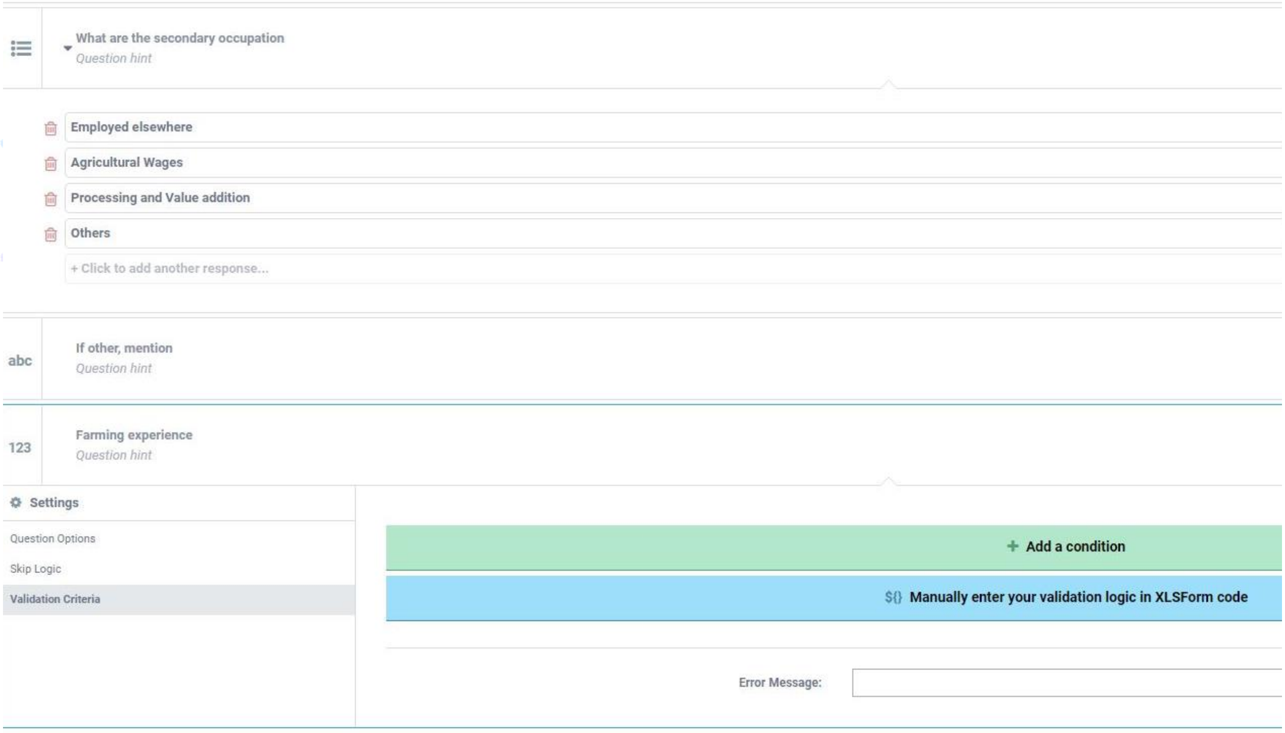
Others

+ Click to add another response...

abc If other, mention

Question hint

- Another useful function is Validation criteria. You can restrict the response values within a limit to avoid data entry errors. For example, the question is “what is the farm gate price of rice in Rs/ Kg?”. The answer has to be less than 50 even by conservative estimates. But, enumerator could get confused and enter 1200 (considering it in Rs/ Quintal). To avoid such errors, we can use validation criteria to limit response to less than 50. If the enumerator enters any value of more than it, error message will pop out, we can even customize the error message to remind him that the price is in Rs/ Kg not in Rs/ Quintal.



What are the secondary occupation
Question hint

- Employed elsewhere
- Agricultural Wages
- Processing and Value addition
- Others

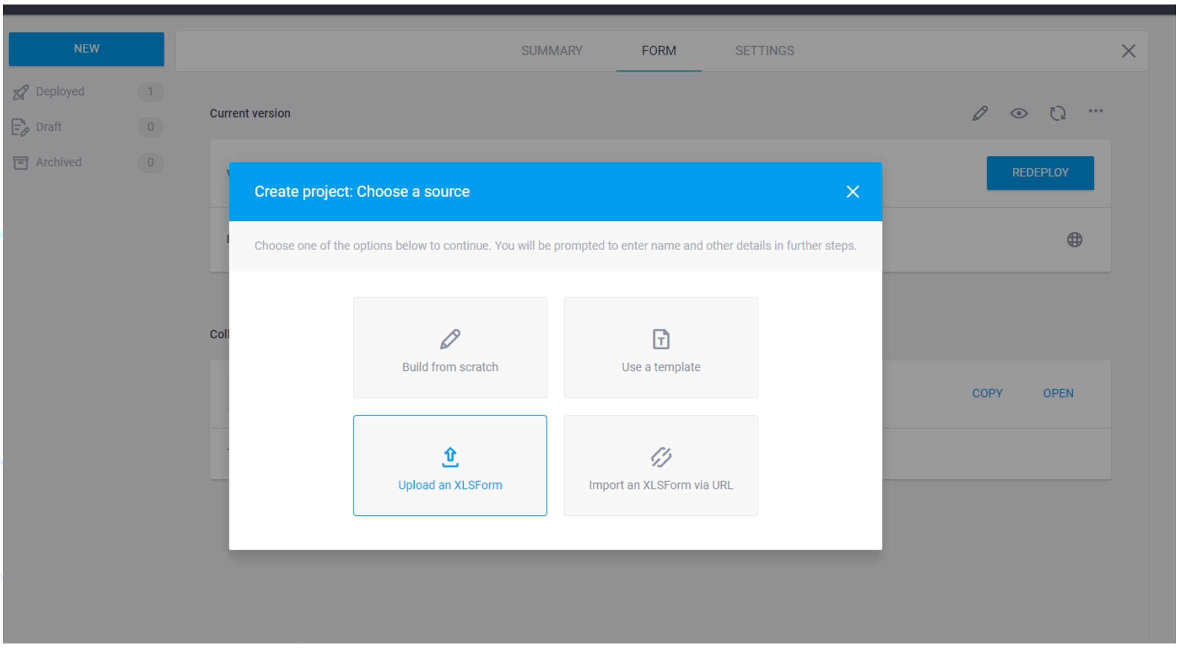
If other, mention
Question hint

Farming experience
Question hint

Error Message:

+ Click to add another response...
abc
123
Settings
Question Options
Skip Logic
Validation Criteria
+ Add a condition
${} Manually enter your validation logic in XLSForm code

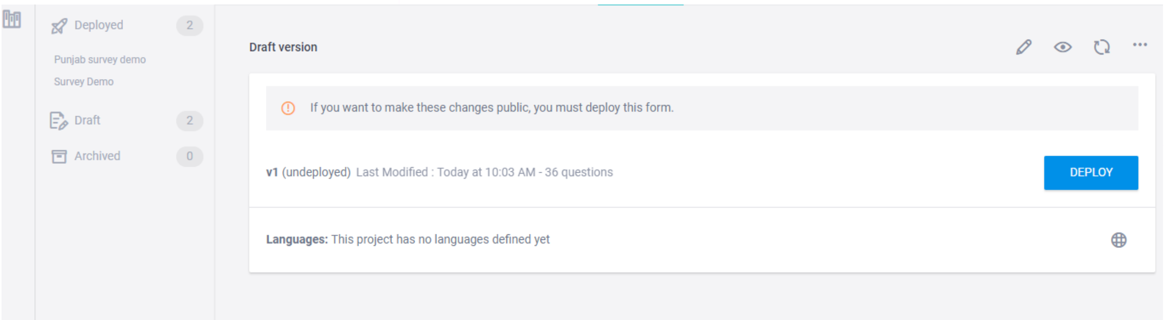
- Making the questionnaire directly in the Kobo toolbox can be time consuming and repetitive. Alternative is to prepare the questionnaire in ‘Open Data Kit (ODK)’ format and upload it directly to kobo. Understanding ODK format at the beginning can be bit tricky, but it will save time in the long run. However, explaining the ODK format is beyond the scope of this blog. One example ODK file is anyhow provided in the link below, which can be directly uploaded to Kobo (link).



Create project: Choose a source

Choose one of the options below to continue. You will be prompted to enter name and other details in further steps.

- Once the questionnaire is ready, first preview it. If satisfied, then deploy the questionnaire.

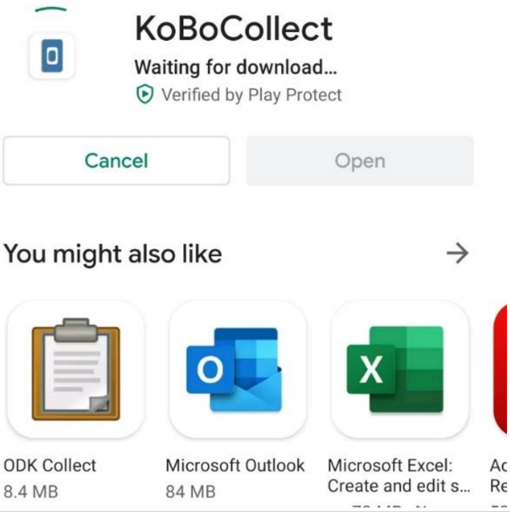


If you want to make these changes public, you must deploy this form.

v1 (undeployed) Last Modified : Today at 10:03 AM - 36 questions

Languages: This project has no languages defined yet

- Step 4: Next step is to download the 'KoBoCollect' in all the devices which will be used for data collection (Preferable Android, in IOS devices on Web Form can work).



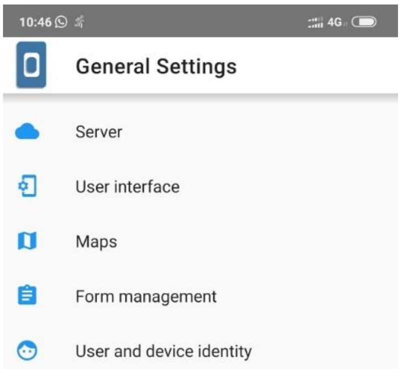
KoBoCollect

Waiting for download...

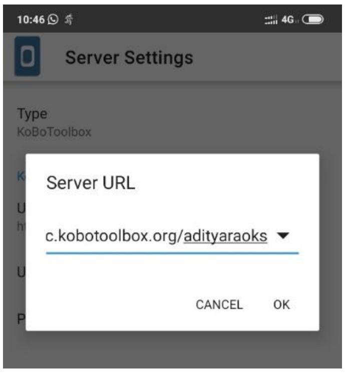
Verified by Play Protect

Step 4

- Step 5: Once installed, go to the general setting- server. Modify the server URL as https://kc.kobotoolbox.org/username - here the in the place of username, you have to input your username used to create the survey schedule. For instance, in the demo survey, the username used is 'adityaraoks'. So the URL is edited to include adityaraoks at the end.



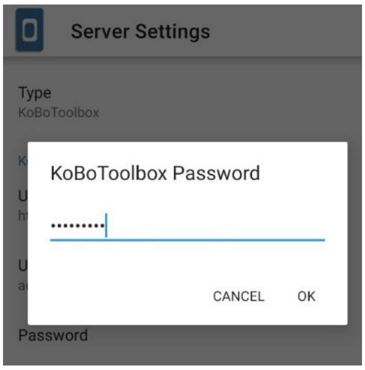
Step 5a



Server URL

c.kobotoolbox.org/adityaraoks

Step 5b



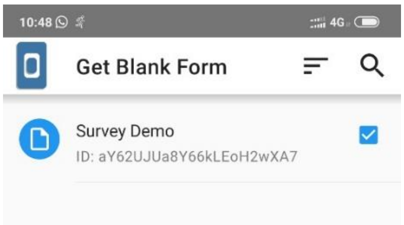
KoBoToolbox Password

•••••••••

Step 6

Step 6: Enter the username and password of the kobo account which is used to generate the survey. This is onetime process and it won't ask for username and password again.

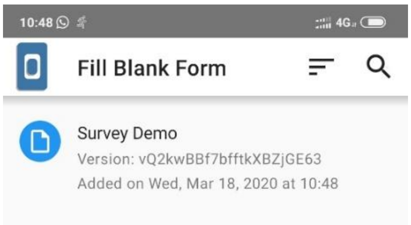
- Step 7: click on the 'Get Blank Form' tab. All the survey's deployed by the respective kobo account will be displayed. Select the survey which you want to fill.
- Step 8: Now go to Fill blank form. Here you will see the questionnaire you have developed.
- Step 9: Now you can see all the questions- answer them till you reach end of the survey
- Step 10: Click on Save and Exit.



Survey Demo ☑

ID: aY62UJUa8Y66kLEoH2wXA7

Step 7

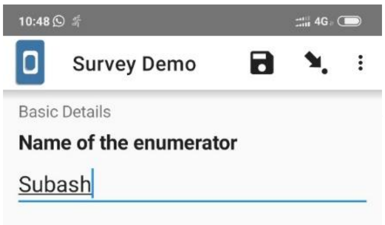


Survey Demo

Version: vQ2kwBBf7bfftkXBZjGE63

Added on Wed, Mar 18, 2020 at 10:48

Step 8

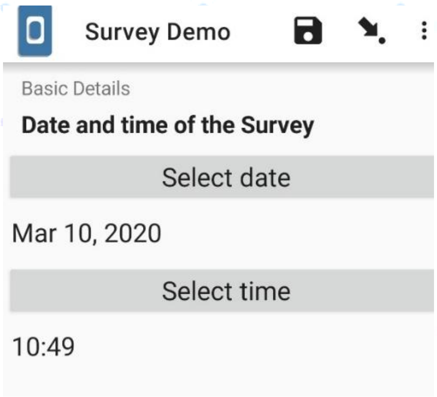


Basic Details

Name of the enumerator

Subash

Step 9a



Survey Demo

Basic Details

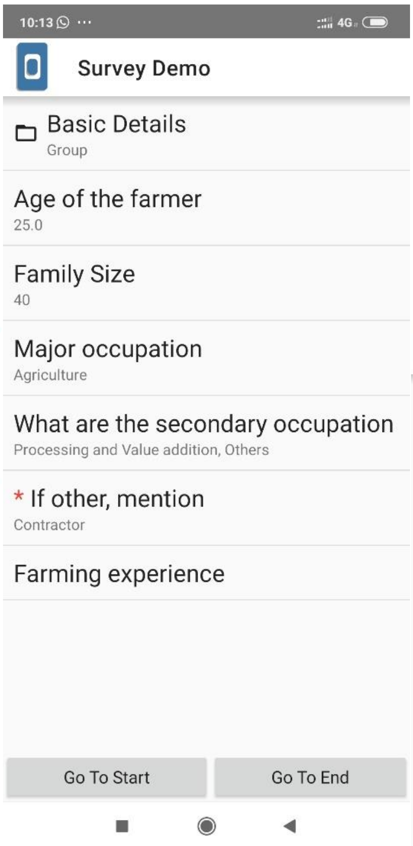
Date and time of the Survey

Select date

Mar 10, 2020

Select time

10:49



10:13

Survey Demo

Basic Details
Group

Age of the farmer
25.0

Family Size
40

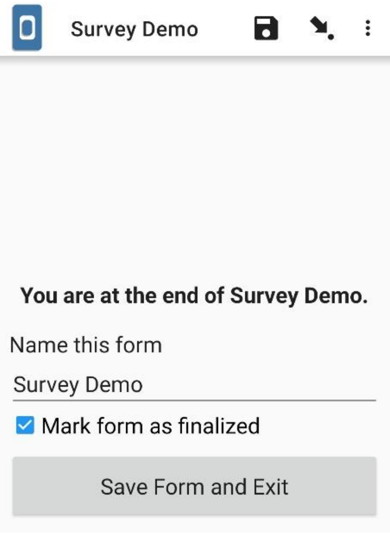
Major occupation
Agriculture

What are the secondary occupation
Processing and Value addition, Others

* If other, mention
Contractor

Farming experience

Go To Start | Go To End



Survey Demo

You are at the end of Survey Demo.

Name this form

Survey Demo

☑ Mark form as finalized

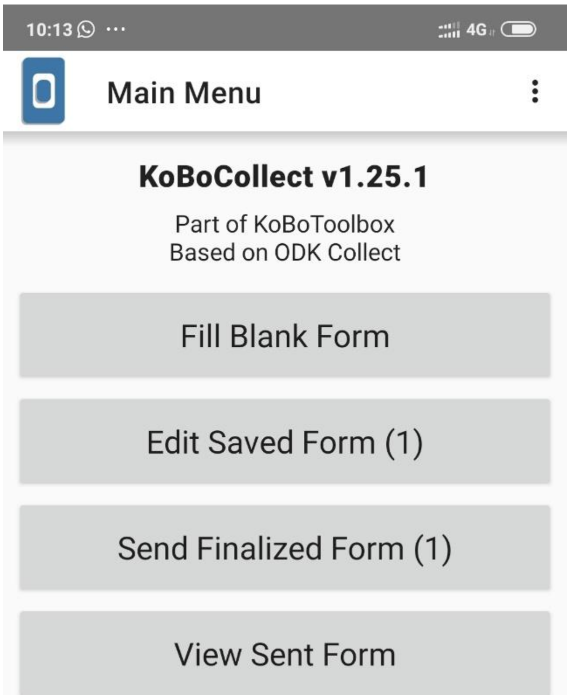
Save Form and Exit

- Step 9b

Step 9c

Step 10

- Step 11: Once you have the internet, click on the send finalized form option. The Survey data collected will be sent to the server, which can be immediately accessed. So, the survey can be conducted even when there is no internet collection. Finished surveys can be sent at the end of the day once you have the internet connection.



10:13

Main Menu

KoBoCollect v1.25.1

Part of KoBoToolbox
Based on ODK Collect

Fill Blank Form

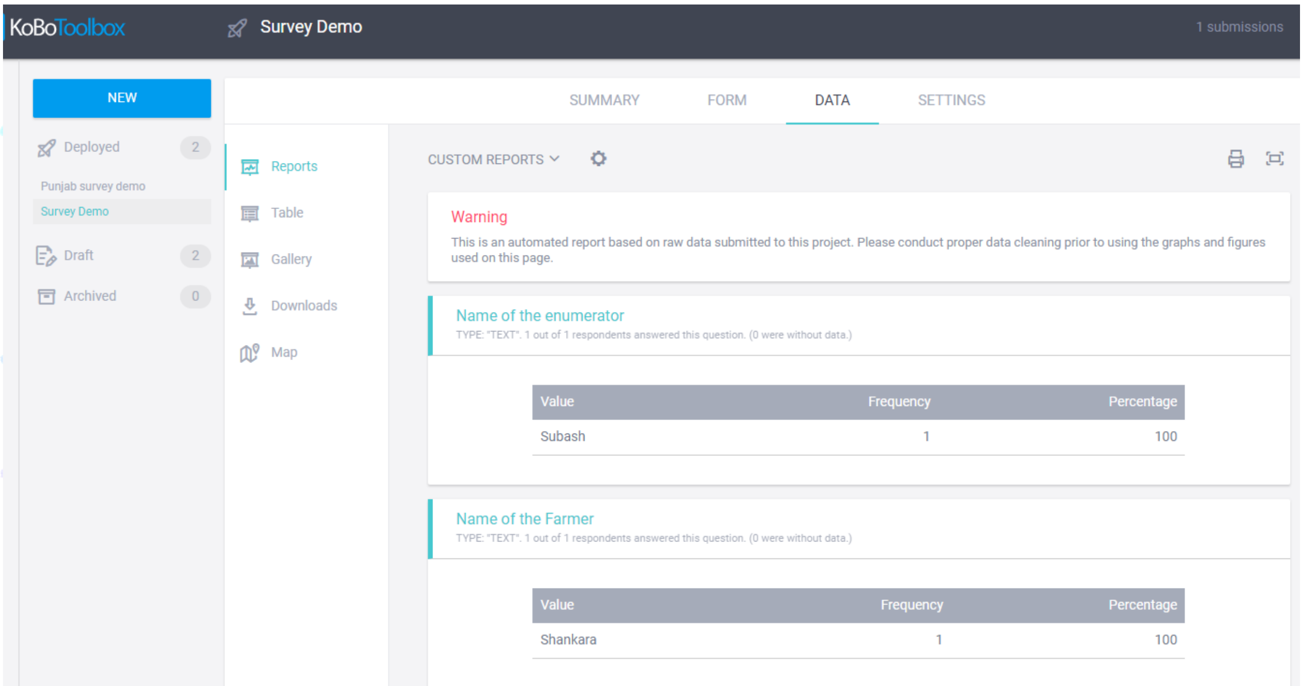
Edit Saved Form (1)

Send Finalized Form (1)

View Sent Form

Step 11

Step 12: Now we will tell you about accessing the collected data from the server. Once the form is sent by the enumerator, we can access the data by clicking on data tab. Custom graphs and tables are displayed here.



Warning

This is an automated report based on raw data submitted to this project. Please conduct proper data cleaning prior to using the graphs and figures used on this page.

Name of the enumerator

TYPE: "TEXT". 1 out of 1 respondents answered this question. (0 were without data.)

Value	Frequency	Percentage
Subash	1	100

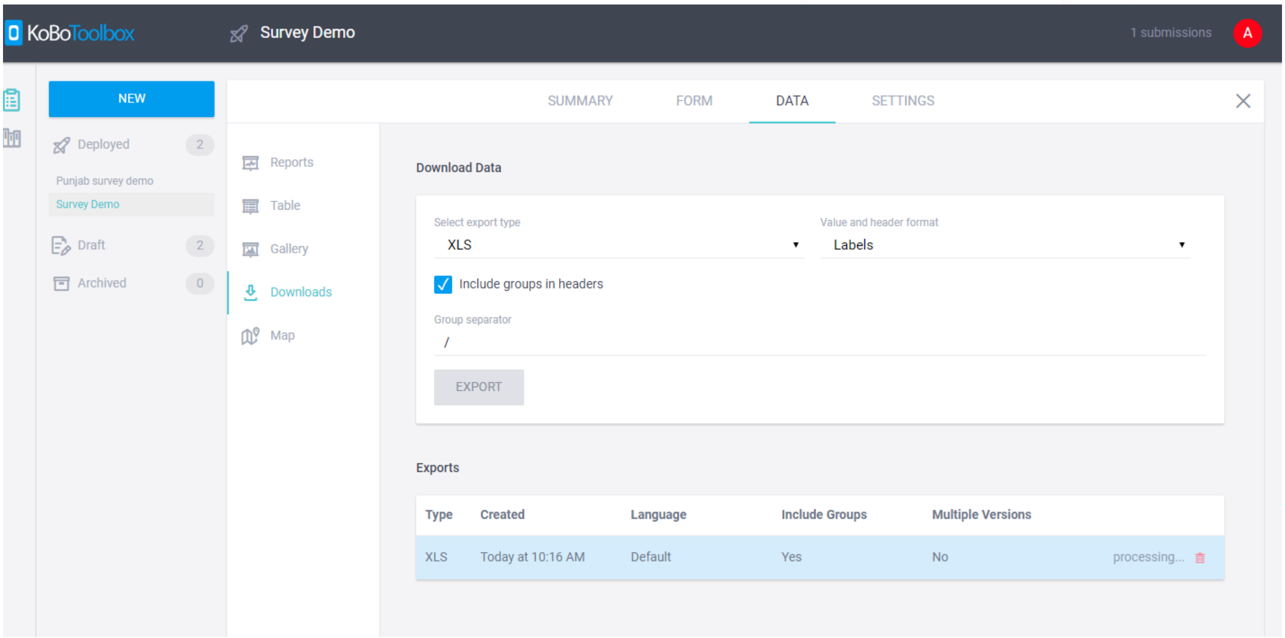
Name of the Farmer

TYPE: "TEXT". 1 out of 1 respondents answered this question. (0 were without data.)

Value	Frequency	Percentage
Shankara	1	100

Step 12

- Step 13: The data can be downloaded in XLS format. The person monitoring the survey can check for the consistency of data by tabulating or summarizing the relevant information. Tabulating by the data by enumerator can also indicate few enumerator specific errors in data entry. Immediate feedback can be given to the respective enumerator about the error.



Download Data

Select export type: XLS

Value and header format: Labels

☑ Include groups in headers

Group separator: /

EXPORT

Exports

Type	Created	Language	Include Groups	Multiple Versions	
XLS	Today at 10:16 AM	Default	Yes	No	processing...

Step 13

Few hints from our personal experience

- Predict the common data collection/ entry mistakes. Even pretesting can be used for this purpose. Then use validation criteria to minimize them.
- If you wish to collect the data in the form of a table, as below

Sl. No. of person	Age	Education	Employment

Customizing table format in Kobo is difficult. It is better to group the questions – Age, Education and Employment and repeat the groups as many times as you wish. You can use skip logic to regulate the number of times the question repeat.

- Pre testing of the questionnaire in CAPI is a must. Sometimes the skip logic may not work or some question may be incorrectly displayed or ordered. The errors can be minimized by identifying them during pre-testing.
- Tabulate the data as frequently as possible during real time monitoring. Tabulate responses by enumerator and observe for patterns in response. For instance, in a question to rank the constraints faced by farmers, if a enumerator follows a typical pattern like 3, 2, 4 and 1 for all the farmers, then there is a possibility that he might be filling responses on his own without even asking. You can talk to the enumerator about it and verify the details.
- On similar lines, another common error is ‘average value of response’. Once an enumerator finish may be 5 or 10 surveys, he gets a vague idea on how much fertilizer is used, what is the seed rate etc. There is a possibility that he starts to enter those average values for all the future respondents. So, closely observe for such errors.

References

Aditya K.S, Subash S.P and Bhuvana N. 2020. Getting smart with Surveys: Using Computers, Tablets or Smart Phones for Household Surveys Agriculture Extension in South Asia (AESA) blog, accessible at https://www.aesanetwork.org/blog-106-getting-smart-with-surveys-using-computers-tablets-or-smart-phones-for-household-surveys/

DIME (2020). Computer-Assisted Personal Interviews (CAPI) https://dimewiki.worldbank.org/wiki/Computer-Assisted_Personal_Interviews_(CAPI)

KoBoToolbox. Available online on https://www.kobotoolbox.org/. Accessed on 19-03-2020.

Designing Visual for Making Effective Presentation

R Roy Burman
Division of Agriculture Extension, ICAR-Indian Agricultural Research Institute, New Delhi

Layout is the arrangement of various visual elements of communication piece — an illustration, a write up (copy, headline, slogan) colour and white space in a pattern that pleases the eye and makes reading easy and convenient.

Before dealing in detail the, various visual elements, we would try and understand what goes in into a good graphic design or layout. For this, let us discuss the principles of layout.

Good graphic design organizes the basic shapes formed by the title, the illustration, and the body copy along with the open space, into an interesting and balanced unit. This basic arrangement should enhance the communication of message, be simple in design and pleasing to the eye.

Some Key Concepts Optical Centre

Many of the underlying principles of design are traceable to the architectural efforts of the ancient Greek's. They still hold true. For instance, the proportion of an object isn't tied only to the rules of mathematics.

You could measure the centre of a page, put your title words there, and still not be satisfied that the title looked centred. Nor will it look pleasing. The human eye has its own illusions. It perceives the centre of a thing slightly above the mathematical centre. We apply this knowledge by placing the words of a cover page above centre so they "look right". For example, we make the bottom marginof a bordered area a bit wider to campmate for the same illusion.

The true or mathematical centre is the point where corner diagonal intersect. A page or an ad however, is not divided at the optical centre a point about five-eighths of the way up the page, the point the eye invariably chooses as the centre of a printed surface.

Hot Spots

There are four hot spots on a page and these are the crossing points when the page is cut on two- thirds from both vertical and horizontal sides.

Centre of Interest

Obviously you need to select one centre of interest and all visual elements cannot compete for equal attention for the reader or viewer. You need to select only one centre of interest and arrange all the other visual elements in such a way that they support the centre of interest in every way. The centre of interest can be an illustration or a body of text or a white space, etc.

Visual Elements

Illustration

It is quite possible to communicate ideas and information by using pictures.

You may need some illustrations - drawings, charts, graphs or pictures - to "dress up" your publication and help get your message across.

Drawings

Tone art for most applications of art in publications, its usually less expensive and more effective to use line art rather than tone art. A black-and-white photographic print would be an example of tone art. Other examples include charcoal, pencil and wash drawings or shaded drawings. They all have varying tones of gray to suggest the structure and shape of the objects in the picture.

Line art, on the other hand, is comprised only of black and white, no grays. For instance, an ink drawing has areas of white lines of black and may be ”hatched" lines to suggest formal changes. Screened photographs are really a form of line art.

Photo sketching can also be done to construct complex illustrations by combing various images. Photo sketching is fastar, cheaper and easier. You can eliminate all distracting elements including the background from the photograph.

Clipart method is another versatile and inexpensive method. Clip art refers to ready-to-use drawings already done by professionals, available in books or on CDs. You can cut them and paste thém as you please on the layout directly.

Realistic picture are of six different styles; out of which there are two styles of photographs and four styles of drawing. Illustration can be the essence of your communication piece or merely cosmetic additions. Both uses are valid. Style and quality are important in either case because, like other graphic elements, the illustrations you choose strongly influence, the viewers’ perceptions of the message.

Typographics

The choices you make about type styles, sizes, case (capitalization), and spacing will also influence how your message is perceived.

Type Styles

The meaning of words are greatly enhanced by right choices of type styles. Type styles can br divided in to two groups: display styles and body types. Display styles are larger are designed tc enhance the meaning of words. The body types are simpler and more legible.

The idea is to catch the attention of your readers and to set your concept quickly, then use ar appropriate display style for the headline, cover page or title. The more legible body style of type fo› the body of the message.

Type Size

Size of the lettering is more important to the design of projected material and display material than it i‹ with printed matter.

Proper letter height for projections/displays is figured on the basis of the distance to the last row of your audience or the distance at which you hope to attract attention to your display. Letters should be 1 inch tall for each 25 feet. That viewing size. The longer sizes for optimum legibility. Gives below ar‹ the letter height sizes and the minimum distance at which it is reasonably visible: Within a message, you may use a variance of sizes to distinguish the move important items from the rest. Emphasis might also be accomplished by using a different colour, style, or case for certain words.

Case

Just as style and size can affect legibility, so can the case. Use both upper-case and lowercase in messages. Words in all uppercase letters are more difficult to read. All capitals words are harder to read because they lack the unique shapes of lowercase.

Spacing

Spacing can also be critical. Letters within a word should be close but not touching. Letters should look as though they have equal space between them, even if the spaces aren't really equal. For instance, some letters may tuck under other letter to look right. Trust your eye. Between words, leave the amount of space needed to insert a lowercase "i" comfortably.

Lines of lettering should have enough space between them (leading) so that the tails (descenders) in the lines below: Leading equal to one-half the height of the capital is usually sufficient. In correct use of space affects the continuity of reading.

Kerning

The space between the letters in a word get automatically adjusted and squeezed-in so that they appear to be compactly together rather than scattered due to the regular space of one lowercase "i".

Typographies

The choices you make about type styles, sizes, case (capitalization)' and spacing will also influence how your message is perceived. The meanings of words are greatly enhanced by the right choices of type styles.

White Space

Watch the professional designers who do makeup for magazines, direct mail, and advertising. Notice how they make strong use of white space. Do the same in your layouts and designs.

We've seen effective advertisements with no headings, no drawings - only extra wide margins and double and triple spaces between paragraphs.

Leave plenty of space between and around most elements, with extra space along the edges of the poster. Avoid the look of crowding. Allow for several fairly large areas of open or unused spaces. A design begins to look crowded whenever the open space areas fall below 20 per cent of the total area. Many successful (layout posters) have 30 to 40 per cent open space.

The margin is the space between your typed area and the edge of the page. Always leave the widest margin at the bottom. Your side and top margins may be alike, or the top margin may be wider. The ratio of the top, side and bottom margin can be 3 : 2 : 2 : 4.

The configuration of open space is just as important to the impact of the basic design as the shapes formed by the lettering and illustration. It's usually desirable to have various sizes and shapes of open space.

Principles of layout

Proportion

Relationship of width to height of the visual's space is based on ideals established by the Greeks. You may have noticed how the eye is attracted by regular shapes such as circles and squares. They hold our attention momentarily. Attractive, yet more interesting is the shape of a rectangle.

The Greeks termed a rectangle 3 unit by 5 units "the Golden Rectangle". They found the proportion to be the most agreeable and pleasing to the eye. We still see these proportions today in 3X5 cards. Given the width of a rectangle, the golden rectangle height can be found by multiplying the width by 1.62. The golden rectangle either vertically or horizontally is as appealing. Keeping good proportion in mind and applying it whenever possible will add greatly to the quality of published work.

Balance

Balance is perhaps the most important principle of layout designing. Balancing is the process of grouping visual elements of a communication piece (parts of a whole) so that they seem to constitute a single unit or order rather than a number of unrelated parts. It's quite undesirable to scatter the visual elements evenly over the entire layout area. At the same time it's usually ineffective, also, to crowd all the elements in to one end or one corner of the area.

Balance is rest or repose. Grouping shapes and masses around the optical centre in such a way that there are equal attractions and weight on either side of the centre obtain this restful effect.

Formal Balance

In formal balance, order is achieved when masses (illustration, headlines, copy, white space) are equally distributed at the right and left of an imaginary vertical line that divides the space in the centre. Formal balance is easy to obtain because one identica[part simply balances with another, as two blocks of copy or illustrations in an advertisement. Such balance, however, is not particularly interesting because it is static. It may be effective and appropriate, however for ads that aim at dignity, conservatism, stability and dependability. This type of symmetrical balance requires very little design ability. Opportunities for alternative arrangement and use of colour for emphasis will be very much limited. Often it may appear very ordinary.

Informal Balance

Informal balance is one in which unequal shapes, weight and colours are placed at such distance from the centre to enable achieve a balance. This balance is asymmetrical and the vertical line does not cut the arrangement of visual elements into two equal halves.

Achieving Balance in a Layout

Balance need not mean having things equal on both the left and right sides of centre (photos opposite photos or words opposite words). Rather, for the sake of visual interest, strive for dynamic balance. Balance the words in a block of copy against white space and a photograph.

Asymmetrical arrangement enable the designer to shift the emphasis of the arrangement away from the centre and allow other areas of interest to be elevated and given prominence while imparting variety to the whole design.

Abstract designs are usually of this type; they enable forms, which vary considerably in size and number to be positioned within the layout, without upsetting the balance. To do this satisfactorily, consideration must be given to "weight" of colour and mass shape.

While balancing a layout, if the units are of the same size and weight, they should be placed at approximately the same distance from the fulcrum, or point of balance. The point is the optical cenfre.

The principle operates like the familiar seesaw. If we have display units of unequal size or weight, the larger or heavier units should be placed nearer the optical centre than the smaller ones.

Balance is necessary to design, but it's often more effective if achieved by means other than by centring the elements. Asymmetrical or informal balance is usually more interesting, more fun to work with, and more challenging. Informal balance is less static and monotonous; it suggests movement.

Don't divide space equally on the page. Use variety to get informal balance. Make a dummy and play moving different shapes and sizes of copy, drawing, and pictures on the drawing board to find a pleasing layout. Make several dummies or thumbnail sketches to see which is more appealing and balanced.

Harmony

The safest way to achieve type harmony is to use are series of type on a design, or throughout the page relying on the light, bold and heavy versions of a single family along with its other variations - i/a//cs, outline and the like.

The scale of grey colour between black and white and between tints and shades, of a colour calls for harmony in tonal values. Careful selection of type, boarders and illustrations provides opportunity tor appropriate tone.

The pleasing relationship in respect to the contour of the parts that go to make up a whole is caller shape harmony. Because the units on a page, headlines, pictures and the like are usually rectangles, rectangles should predominate.

Contrast

Contrast refers to a form of emphasis to make things stand out by the use of striking comparison. Contrast in a sense is the opposite of harmony

Contrast is a means of directing reader's interest on the page. A dark place on a light page attracts interest, as does a light spot on a predominantly dark page. Large type amidst a page full of small type dominates the page. The same sort of emphasis is given to small type when its contrasted against a summoning area of while space.

A warning is in order about false reasoning in the use of contrast. It may seem that if modest contrast help to "give an attractive appearance", strong contrast will do just that much better. Various styles of typefaces can bring contrast to type pages of a layout where copy is the centre of interest.

Repetition

Repeating a line, curve, shape or texture also produces interesting designs. Repeat a cover picture, or a logo, or an emblem or a drawing or a theme on the inside pages for effectiveness and reinforcement of the theme.

Order

Order is a goal. There is nothing vague or random about a good graphic design. The element are arranged in a way that movement is rhythmic and not haphazard. Design should be rhythmic. Every element should be arranged to lead the eye progressively from one part to the next, enhancing the reader's comprehension until every part is seen. If any element distracts the reader or disrupts the **orderly comprehension of the message in a design, then that is a bad design.**

Variety

Variety refers to the diversity of related parts or elements in a design. The masses covered by copy, drawing and title should not be identical. They should be of different shapes and sizes. It is the dash of spice, which relieves monotony.

Unity

All the visual elements in the design should look like they really belong to one another. There should be unity among the parts of the design in communicating a message.

Emphasis

It means stress upon an element within a design. By emphasis the eye is carried first to the most important thing in any layout and from that part to every other detail in their order of importance. In a superior design, either the illustration or the lettering dominates, rather than an equal division between the two. A design with the illustration occupying more area then the lettering grabs a lot of attention, but you must take care to assure the illustration is supponive of the message.

Focus

It is often desirable to develop a strong centre of interest on the words or the illustration that is the quickly grasped key to your message. This reinforces the basic task that you set out to accomplish in the first place. Avoid placing the centre of interest in the geometrical centre of the poster area. Also, refrain from placing it too close to the edge of the design, or tightly crammed in a corner. Roughly, a third of the distance up, down, or in from the edge is much more desirable.

Artificial Neural Networks: An Introduction

Girish Kumar Jha
Division of Agricultural Economics, ICAR-Indian Agricultural Research Institute, New Delhi

Introduction

The seed of the modern era of neural networks was sown with the pioneering work of McCulloch and Pitts in 1943 when they introduced the idea of neural networks as computing machines. The major impetus to the growth and development of research and applications in the area of neural networks was provided by the development of the back-propagation algorithm, a popular learning algorithm for the training of multilayer perceptrons by Rumelhart, Hinton and Williams (1986). However, the recent resurgence of interest in this area is mainly because neural networks are the fundamental building block of deep learning which is an artificial intelligence function that allows computational models to learn features from raw data with multiple levels of abstraction. Presently, artificial intelligence (AI) is a growing field and has many practical applications due to the availability of massive data, graphics processing units (GPUs) hardware and open source software like python and tensorflow etc.

Artificial neural networks (ANNs) are non-linear data driven self-adaptive approach as opposed to the traditional model-based methods. They are powerful tools for modelling, especially when the underlying data relationship is unknown. ANNs can identify and learn correlated patterns between input data sets and corresponding target values. After training, ANNs can be used to predict the outcome of new independent input data. ANNs imitate the learning process of the human brain and can process problems involving non-linear and complex data even if the data are imprecise and noisy. Thus, they are ideally suited for the modeling of agricultural data which are known to be complex and often non-linear. In recent years neural computing has emerged as a practical technology, with successful applications in many fields as diverse as finance, medicine, engineering, geology, physics and biology. The excitement stems from the fact that these networks are attempts to model the capabilities of the human brain. From a statistical perspective neural networks are interesting because of their potential use in prediction and classification problems.

A very important feature of these networks is their adaptive nature, where “learning by example” replaces “programming” in solving problems. This feature makes such computational models very appealing in application domains where one has little or incomplete understanding of the problem to be solved but where training data is readily available. These networks are “neural” in the sense that they may have been inspired by neuroscience but not necessarily because they are faithful models of biological neural or cognitive phenomena. In fact, majority of the network are more closely related to traditional mathematical and/or statistical models such as non-parametric pattern classifiers, clustering algorithms, nonlinear filters, and statistical regression models than they are to neurobiology models.

Neural networks (NNs) have been used for a wide variety of applications where statistical methods are traditionally employed. They have been used in classification problems, such as identifying underwater sonar currents, recognizing speech, and predicting the secondary structure of globular proteins. In time-series applications, NNs have been used in predicting stock market performance. As statisticians or users of statistics, these problems are normally solved through classical statistical methods, such as discriminant analysis, logistic regression, Bayes analysis, multiple regression, and ARIMA time-series models. It is, therefore, time to recognize neural networks as a powerful tool for data analysis.

Basics of a neuron

An artificial neural network is a set of simple computational units that are highly interconnected. The units are also called nodes and loosely represent the biological neuron. A graphical presentation of neuron is given in Figure 1. A neuron is an information processing unit that is fundamental to the operation of a neural network. The connections between nodes are unidirectional and are represented by arrows in the figure. These connections model the synaptic connections in the brain. Each connection has a weight called the synaptic weight, denoted as w_{kj}, associated with it. The synaptic weight, w_{kj}, is interpreted as the strength of the connection from the *j*th unit to the *k*th unit. Unlike a synapse in the brain, the synaptic weight of an artificial neuron may lie in a range that includes negative as well as positive values. If a weight is negative, it is termed inhibitory because it decreases the net input. If the weight is positive, the contribution is excitatory because it increases the net input.

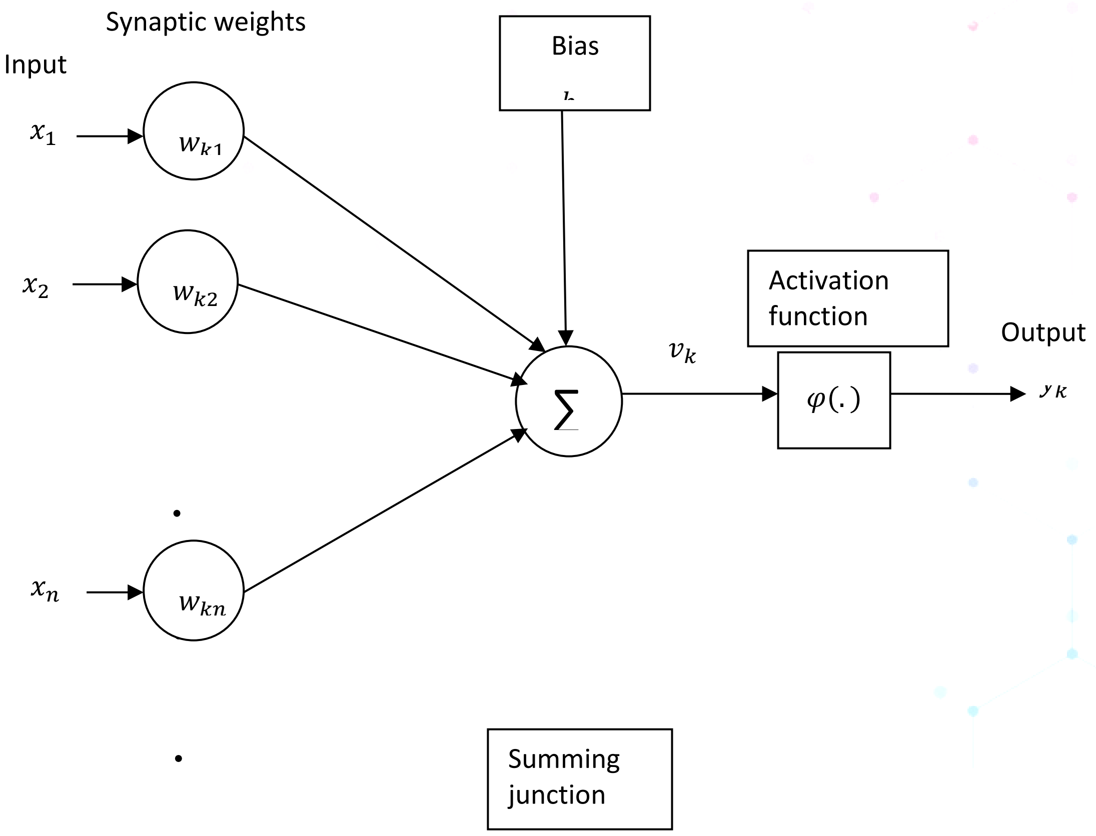


Figure 1: Nonlinear model of a neuron

The input into a node is a weighted sum of the outputs from nodes connected to it. Each unit takes its net input and applies an activation function to it. The neuronal model of Figure 1 also includes an externally applied bias, denoted by b_k. The bias b_k has the effect of increasing or lowering the net input of the activation function depending on whether it is positive or negative respectively. In mathematical terms, we may describe a neuron *k* by the following equations

$$y_k = \varphi(v_k) = \varphi\left(\sum_{j=1}^{n} w_{kj}x_j + b_k\right)$$

where $x_1, x_2, \ . \ . \ . \ , x_n$ are the input patterns, $w_{k1}, w_{k2}, \ . \ . \ . \ , w_{kn}$ are the synaptic weights of neuron k, b_k is the bias, $\varphi(.)$ is the activation function and y_k is the output of the neuron. The neural networks are built from layers of neurons connected so that one layer receives input from the preceding layer of neurons and passes the output on to the subsequent layer.

Types of activation function

An activation function which is also known as squashing function, squashes or limits the amplitude range of the output of a neuron. It is a mathematical function which converts the input to an output, and adds the magic of neural network processing. The abstraction of the processing of neural networks is mainly achieved through the activation functions. Activation functions give the nonlinearity property to neural networks and make them true universal function approximators. Three commonly used activation functions are described below:

a. **Sigmoid Function**: The term sigmoid means S-shaped and the logistic form of the sigmoid maps the interval (-∞, ∞) onto (0, 1). The main motivation of using this activation function is allowing the outputs to be given a probabilistic interpretation. It is defined by

$$f(x) = \frac{1}{(1 + e^{-ax})}$$

where *a* is the slope parameter of the sigmoid function and is illustrated in Figure 2.

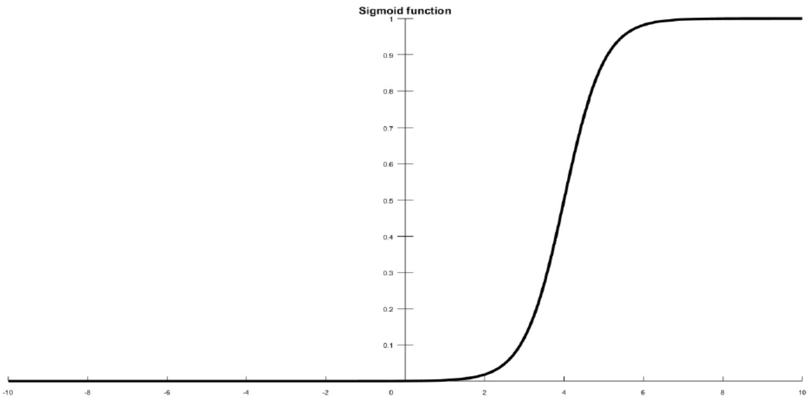


Figure 2: Sigmoid function

b. **Hyperbolic Tangent**: This is a nonlinear function, defined in the range of values (-1, 1) and is plotted in

Figure 3. This function is defined by

$$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

It is established empirically that tanh function provided faster convergence of training algorithms than logistic function. Both the logistic and hyperbolic tangent functions differ only through a linear transformation. These two were the most common form of activation functions used in the construction of neural networks prior to the introduction of rectified linear units.

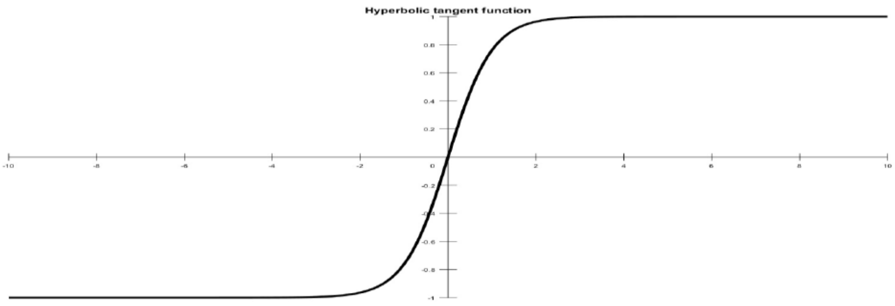


Figure 3: Hyperbolic tangent function

c. **Rectified Linear Unit (ReLU):** It is the most used activation function since 2015. It is a simple condition and has advantages over the other functions. The function is defined by the following formula and is plotted in Figure 4:

$$f(x) = \begin{cases} 0 & when\ x < 0 \\ x & when\ x \geq 0 \end{cases}$$

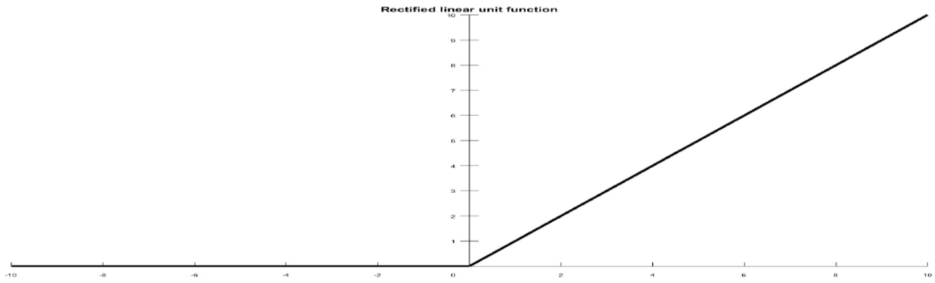


Figure 4: Rectified Linear Unit

Neural networks architectures

An artificial neural network is defined as a data processing system consisting of a large number of simple highly inter connected processing elements (artificial neurons) in an architecture inspired by the structure of the cerebral cortex of the brain. There are several types of architecture of neural networks. However, the two most widely used ANNs are discussed below:

Feed forward networks

In a feed forward network, information flows in one direction along connecting pathways, from the input layer via the hidden layers to the final output layer. There is no feedback (loops) i.e., the output of any layer does not affect that same or preceding layer.

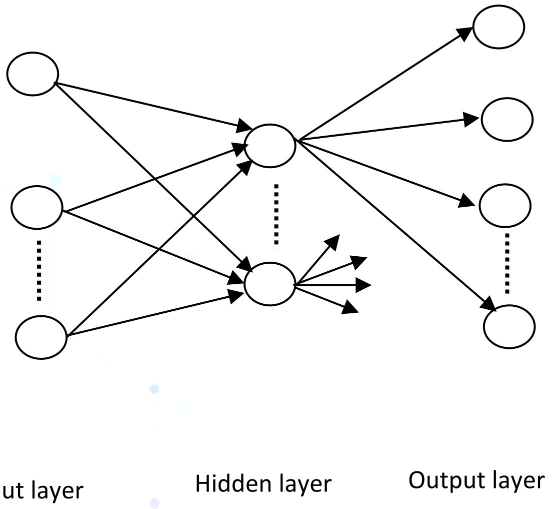


Figure 5: A multi-layer feed forward neural network

Recurrent networks

These networks differ from feed forward network architectures in the sense that there is at least one feedback loop. Thus, in these networks, for example, there could exist one layer with feedback connections as shown in figure below. There could also be neurons with self-feedback links, i.e. the output of a neuron is fed back into itself as input.

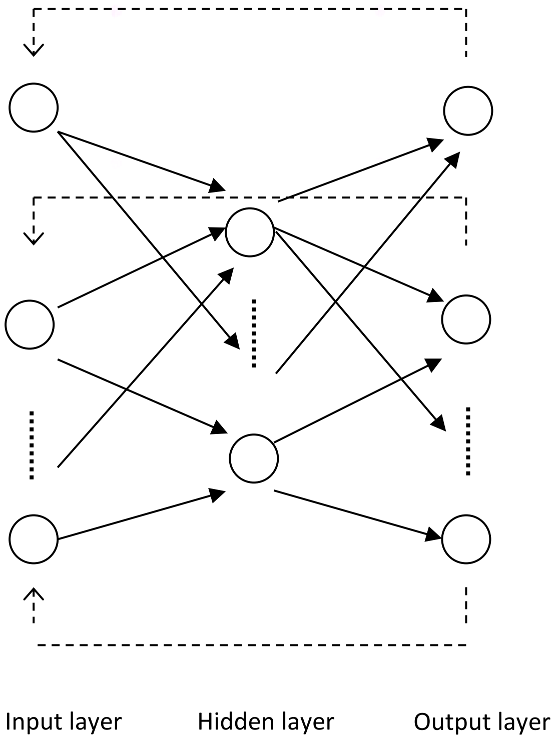


Figure 6: A recurrent neural network

Learning/Training methods

Learning methods in neural networks can be broadly classified into three basic types: supervised, unsupervised and reinforced.

Supervised learning

In this, every input pattern that is used to train the network is associated with an output pattern, which is the target or the desired pattern. A teacher is assumed to be present during the learning process, when a comparison is made between the network's computed output and the correct expected output, to determine the error. The error can then be used to change network parameters, which result in an improvement in performance.

Unsupervised learning

In this learning method, the target output is not presented to the network. It is as if there is no teacher to present the desired patterns and hence, the system learns of its own by discovering and adapting to structural features in the input patterns.

Reinforced learning

In this method, a teacher though available, does not present the expected answer but only indicates if the computed output is correct or incorrect. The information provided helps the network in its learning process. A reward is given for a correct answer computed and a penalty for a wrong answer. Reinforced learning was not one of the popular forms of learning but gaining importance in case of deep learning.

Development of an ANN model

The various steps in developing a neural network model are:

A. Variable selection

The input variables important for modeling variable(s) under study are selected by suitable variable selection procedures.

B. Formation of training, testing and validation sets

The data set is divided into three distinct sets called training, testing and validation sets. The training set is the largest set and is used by neural network to learn patterns present in the data. The testing set is used to evaluate the generalization ability of a supposedly trained network. A final check on the performance of the trained network is made using validation set.

C. Neural network architecture

Neural network architecture defines its structure including number of hidden layers, number of hidden nodes and number of output nodes etc.

- Number of hidden layers: The hidden layer(s) provide the network with its ability to generalize. In theory, a neural network with one hidden layer with a sufficient number of hidden neurons is capable of approximating any continuous function. In practice, neural network with one and occasionally two hidden layers are widely used and have to perform very well.

- Number of hidden nodes: There is no magic formula for selecting the optimum number of hidden neurons. However, some thumb rules are available for calculating number of hidden neurons like for a three layers network with n input and m output neurons, the hidden layer would have *sqrt(n*m)* neurons.
- Number of output nodes: Neural networks with multiple outputs, especially if these outputs are widely spaced, will produce inferior results as compared to a network with a single output.
- Activation function: As mentioned earlier, activation functions are mathematical formulae that determine the output of a processing node. Each unit takes its net input and applies an activation function to it. The purpose of the transfer function is to prevent output from reaching very large value which can 'paralyze' neural networks and thereby inhibit training. Now the default recommendation is to use rectified linear units as activation function for network learning.

D. Evaluation criteria

For regression problems, the sum of squares error function is commonly used in neural networks. The cross-entropy loss function is commonly used in case of classification problem.

E. Neural network training

Training a neural network to learn patterns in the data involves iteratively presenting it with examples of the correct known answers. The objective of training is to find the set of weights between the neurons that determine the global minimum of error function. This involves decision regarding the number of iterations i.e., when to stop training a neural network and the selection of learning rate (a constant of proportionality which determines the size of the weight adjustments made at each iteration) and momentum values (how past weight changes affect current weight changes).

Conclusion

The computing world has a lot to gain from neural networks. Their ability to learn by example makes them very flexible and powerful. A large number of claims have been made about the modeling capabilities of neural networks, some exaggerated and some justified. Hence, to best utilize ANNs for different problems, it is essential to understand the potential as well as limitations of neural networks. For some tasks, neural networks will never replace conventional methods, but for a growing list of applications, the neural architecture will provide either an alternative or a complement to these existing techniques. Finally, I would like to conclude that the performance of traditional/shallow neural networks depend on the feature of the data provided by the domain experts. This limitation inspired another subset of neural networks called deep neural networks (DNNs). DNNs extend traditional ANNs by adding multiple processing layers between input and output layers into the model that allows hierarchical representation of raw data through several layers of abstraction.

Practical Exercise using R Software

The details of construction and training of a time-delay neural network for predictive modeling using R software will be illustrated with the help of example during the webinar.

References

Cheng, B. and Titterington, D. M. (1994). Neural networks: A review from a statistical perspective. Statistical Science, 9, 2-54.

Jha, G.K., Thulasiraman, P. and Thulasiram, R. K. (2009). PSO based neural network for time series forecasting. In proceeding of the IEEE International Joint Conference on Neural Networks, USA, pp 1422-1427.

Jha, G.K. and Sinha, K. (2012). Time-delay neural networks for time series prediction: an application to the monthly wholesale price of oilseeds in India. Neural Comput & Applic, DOI 10.1007/s00521-012-1264-z.

*Kaastra, I. and Boyd, M. (1996). Designing a neural network for forecasting financial and economic time series. Neurocomputing, **10**, 215-236.*

*Kohzadi, N., Boyd, S.M., Kermanshahi, B. and Kaastra, I. (1996). A comparision of artificial neural network and time series models for forecasting commodity prices. Neurocomputing, **10**, 169-181.*

McCulloch, W.S. and Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. Bulletin of Mathematical Biophysics, 5, 115-133.

Patterson, D. (1996). Artificial Neural Networks. Singapore: Prentice Hall.

*Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage ang organization in the brain. Psychological review, **65,** 386-408.*

Rumelhart, D.E., Hinton, G.E. and Williams, R.J. (1986). Learning internal representation by error propagation. In D.E. Rumelhart, J.L. McClelland and the PDP Research Group (Eds.), Parallel distributed processing: Exploration in microstructure of cognition, Volume 1: Foundations, pp 318-362. Cambridge, MA: MIT Press.

Simon Haykin (2006). Neural Networks: A comprehensive foundation, Pearson Prentice Hall.

*Saanzogni, Louis and Kerr, Don (2001) Milk production estimate using feed forward artificial neural networks. Computer and Electronics in Agriculture, **32**, 21-30.*

Swanson, N. R., and White, H. (1997). Forecasting economic time series using adaptive versus non-adaptive and linear versus nonlinear economic models. International Journal of Forecasting, 13, 439–461.

*Warner, B. and Misra, M. (1996). Understanding neural networks as statistical tools. American Statistician, **50**, 284-93.*

*Zhang, G., Patuwo, B. E. and Hu, M. Y. (1998). Forecasting with artificial neural networks: The state of the art. International Journal of Forecasting, **14**, 35-62.*

*Zhang, G.P. (2007). Avoiding pitfalls in neural network research. IEEE transactions on systems, man and cybernetics-Part C: Applications and reviews, **37**, 3-16.*

Choice Analysis: Basics of Conjoint Analysis in Quantifying Attribute Preference

P Venkatesh and Praveen K.V.
Division of Agricultural Economics ,ICAR-Indian Agricultural Research Institute, New Delhi

Introduction

Consumer behavioural analysis is the most important requirement for investors or entrepreneurs. Whenever the entrepreneur starts a new business or introducing a new product in the market, the first requirement is what is the demand for the product and what are all the quality attributes demanded by the consumers. The choice analysis methods are very useful for identifying attributes or traits of the products. Broadly there are two types of choice analysis namely, revealed preference (RP) techniques (eg.travel cost method, hedonic pricing analysis) and stated preference (SP) techniques (eg. contingent valuation methods, choice experiments, conjoint analysis). The RP techniques are based on the actual choice of the consumers, conversely, SP techniques are based on the hypothetical scenarios. For example, in a mobile shop, there are a number of brands (or same brand) with various attributes such as RAM, operating system, storage, camera quality and price. When we collect the sales data of mobiles along with it attributes and analyse and identifying the most preferred attributes then it is a type of revealed preference techniques, where we have observed actual purchasing behaviour of the consumers. Whereas, if we conduct the consumer survey with a hypothetical scenario of different attributes of mobiles phones and elicit the most preferred attributes of mobile phones, then it is a type of stated preference techniques.

Conjoint analysis

Conjoint analysis (CA) is one of the stated preference methods. It is used to understand the how consumers make complex choices among the various alternatives which has trade-offs. For example, one chooses low price mobile phones, then he (she) has to compromise OS and RAM etc. Everyone makes choices in day to day life like purchase of dress, mobiles, choosing restaurants for dinner etc which all involves mental conjoint analysis that contains multiple elements that lead us to our choice. CA is based on theory of demand (consumers derives utility or value from the attributes of the product) and theory of random utility (stochastic preference i.e. consumer may choose different choice from the same subset of alternatives at repeated presentation). Consumer choice decisions are based on the intrinsic and extrinsic cues of the product. The intrinsic cues are part of the physical properties of the product (eg. RAM. OS of mobile phone) and extrinsic cues are not part of the physical properties of the product (eg. price and brand). The traditional ranking method or rating survey cannot place the value for the attributes of the product. On the other hand, CA used to determine consumers preference by conjointly analysing their trade-offs between attributes. One of the advantages of the CA is that it provides relative importance of each attributes of the product (Lee et al, 2015).

Examples of conjoint analysis

CA mostly applied in market research analysis where to capture the consumer choice or preferences. Some of the examples for application of CA are as follows.

- Consumers preference for house: Consumers will be interested in location of the house like near to school, market, railway station, bus stations, hospital, size of the house, and other amenities and reprice of the house. Some consumers will be concentrated on price and some may be having preference for amnesties and some may be for locational advantage. The CA will identify the most preferred combination of attributes of the house as well as the importance of each attributes.
- Farmers preference for a variety: A variety may have various attributes like, duration, drought tolerance, pest and disease resistance, suitable for rainfed, yield, price of the seed, premium price in the market etc. The breeder cannot bring all the best attributes in single variety. Hence, he (she) wants to prioritize the breeding programme based on the farmers preference /need. CA can be useful tool to identify the most preferred attributes of the variety.

Steps in conjoint analysis

We will discuss the study on identification of preferred varietal attributes of pigeonpea variety by using hypothetical datasets.

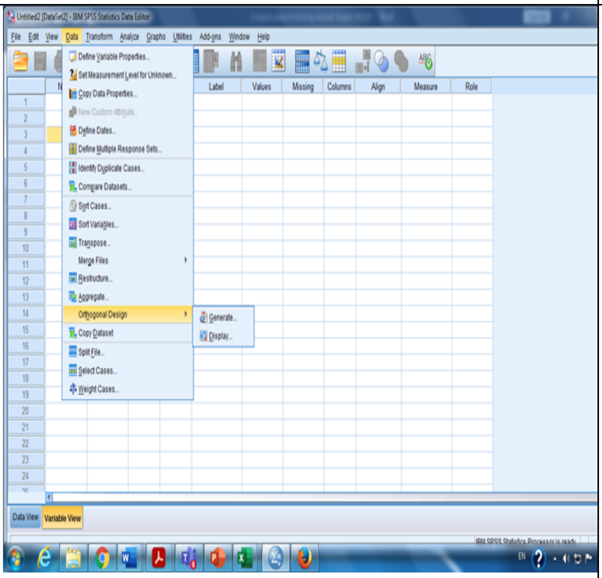
1. **Identification of attributes and their levels:** Attributes are characteristics of the variety for example yield and it has three levels 10-15 q, 15-20 q and > 20 q. Similarly, other attributes ae and their levels are given below.

Sl. No.	Attributes	Levels		
1	Drought	Moderate resistant	Highly resistant	
2	Pod borer	Moderate resistant	Highly resistant	
3	Height	Short	Long	
4	Duration	Short (130 days)	Medium (13-150days)	Long (>150 days)
5	Yield (q/ha)	10-15 q	15 -20 q	> 20 q

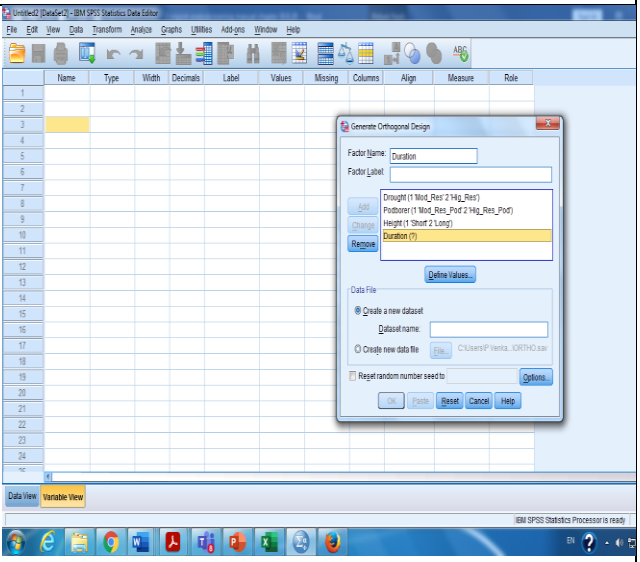
2. **Preparation of orthogonal design:** We have considered five attributes and each attribute is having different levels. In total (2x2x2x3x3=72) all possible combinations (sets) will be formed. However, it will be difficult for the respondents (farmers) to rank all these combinations. Therefore, by using orthogonal design, we can prepare a manageable number of combinations. Orthogonal Design procedure creates a reduced set of varietal combinations that is small enough to include in a survey but large enough to assess the relative importance of each attributes. By using SPSS, we can generate orthogonal design and plan file will be generated.
3. **Data collection:** A survey will be conducted among the farmers to rank the chosen level of combinations.
4. **Data analysis:** By using SPSS data can be analysed. Both plan file and data files are required for the analysis.
5. **Results:** SPSS will produce both utility files as well relative importance of the factors. By using utility values of each attributes, we can estimate the total utility value s of each combinations and we can find the most proffered combinations.

Analytical procedure in SPSS

1.Orthogonal design create



2.Defining factors


Untitled2 [DataSet2] - IBM SPSS Statistics Data Editor
Generate Orthogonal Design
Add
Change
Remove
Define Values...
Options...
OK
Paste
Reset
Cancel
Help
IBM SPSS Statistics Processor is ready

Factor Name: Duration

Factor Label:

Drought (1 'Mod_Res' 2 'Hig_Res')
Podborer (1 'Mod_Res_Pod' 2 'Hig_Res_Pod')
Height (1 'Short' 2 'Long')
Duration (?)

Data File

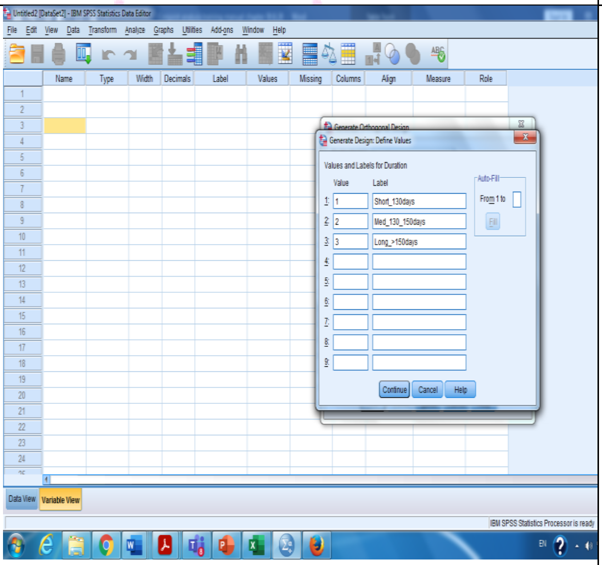
● Create a new dataset

Dataset name:

○ Create new data file C:\Users\P Venka...\ORTHO.sav

□ Reset random number seed to

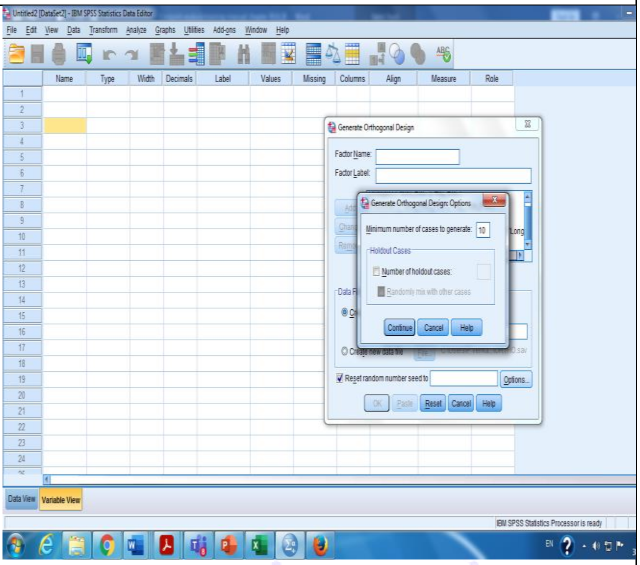
3. Defining factor levels


Untitled2 [DataSet2] - IBM SPSS Statistics Data Editor
Generate Orthogonal Design
Generate Design: Define Values
Auto-Fill
From 1 to
Fill
Continue
Cancel
Help
IBM SPSS Statistics Processor is ready

Values and Labels for Duration

	Value	Label
1:	1	Short_130days
2:	2	Med_130_150days
3:	3	Long_>150days
4:		
5:		
6:		
7:		
8:		
9:		

4.Setting minimum number of cases


Untitled2 [DataSet2] - IBM SPSS Statistics Data Editor
Generate Orthogonal Design
Generate Orthogonal Design: Options
Continue
Cancel
Help
Options...
OK
Paste
Reset
Cancel
Help
IBM SPSS Statistics Processor is ready

Minimum number of cases to generate: 10

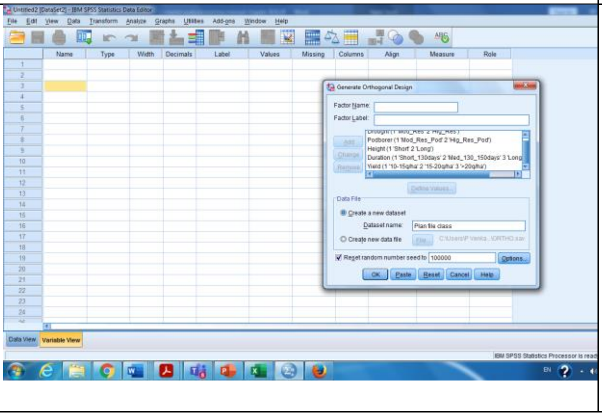
Holdout Cases

□ Number of holdout cases:

Randomly mix with other cases

☑ Reset random number seed to

5.Setting seed value and defining file name


Untitled2 [DataSet2] - IBM SPSS Statistics Data Editor
Generate Orthogonal Design
Add
Change
Remove
Define Values...
Options...
OK
Paste
Reset
Cancel
Help
IBM SPSS Statistics Processor is ready

Factor Name:

Factor Label:

Drought (1 'Mod_Res' 2 'Hig_Res')
Podborer (1 'Mod_Res_Pod' 2 'Hig_Res_Pod')
Height (1 'Short' 2 'Long')
Duration (1 'Short_130days' 2 'Med_130_150days' 3 'Long
Yield (1 '10-15q/ha' 2 '15-20q/ha' 3 '>20q/ha')

Data File

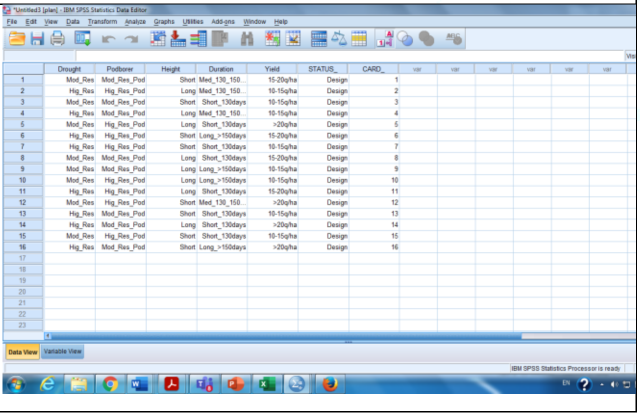
● Create a new dataset

Dataset name: Plan file class

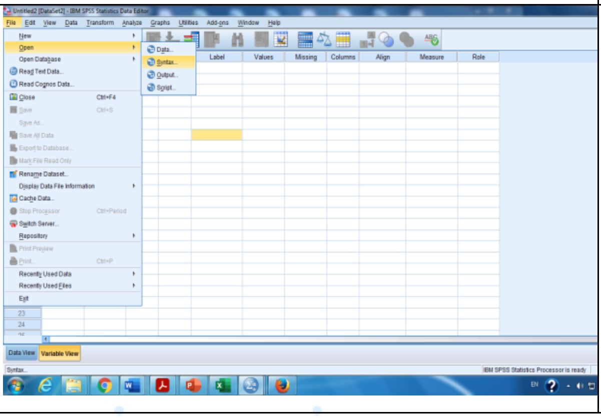
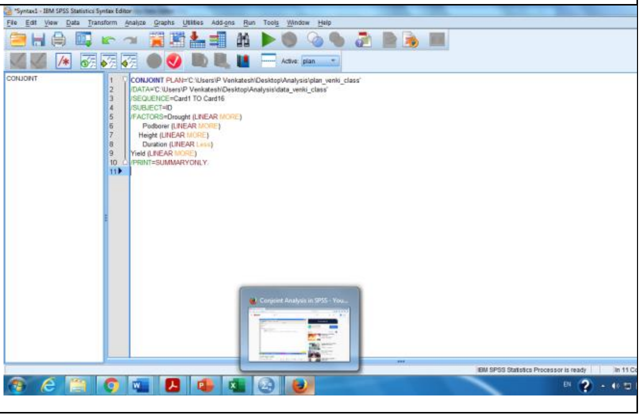
○ Create new data file C:\Users\P Venka...\ORTHO.sav

☑ Reset random number seed to 100000

6.Plan file generated


*Untitled3 [plan] - IBM SPSS Statistics Data Editor
File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help
Data View
Variable View
IBM SPSS Statistics Processor is ready

	Drought	Podborer	Height	Duration	Yield	STATUS_	CARD_	var
1	Mod_Res	Mod_Res_Pod	Short	Med_130_150...	15-20q/ha	Design	1	
2	Hig_Res	Hig_Res_Pod	Long	Med_130_150...	10-15q/ha	Design	2	
3	Mod_Res	Mod_Res_Pod	Short	Short_130days	10-15q/ha	Design	3	
4	Hig_Res	Mod_Res_Pod	Long	Med_130_150...	10-15q/ha	Design	4	
5	Mod_Res	Hig_Res_Pod	Long	Short_130days	>20q/ha	Design	5	
6	Hig_Res	Hig_Res_Pod	Short	Long_>150days	15-20q/ha	Design	6	
7	Hig_Res	Hig_Res_Pod	Short	Short_130days	10-15q/ha	Design	7	
8	Mod_Res	Mod_Res_Pod	Long	Short_130days	15-20q/ha	Design	8	
9	Mod_Res	Mod_Res_Pod	Long	Long_>150days	10-15q/ha	Design	9	
10	Mod_Res	Hig_Res_Pod	Long	Long_>150days	10-15q/ha	Design	10	
11	Hig_Res	Hig_Res_Pod	Long	Short_130days	15-20q/ha	Design	11	
12	Mod_Res	Hig_Res_Pod	Short	Med_130_150...	>20q/ha	Design	12	
13	Hig_Res	Mod_Res_Pod	Short	Short_130days	10-15q/ha	Design	13	
14	Hig_Res	Mod_Res_Pod	Long	Short_130days	>20q/ha	Design	14	
15	Mod_Res	Hig_Res_Pod	Short	Short_130days	10-15q/ha	Design	15	
16	Hig_Res	Mod_Res_Pod	Short	Long_>150days	>20q/ha	Design	16	

7.Opening of a syntax file	8.Syntax for analysis
	

*Syntax1 - IBM SPSS Statistics Syntax Editor
File
Edit
View
Data
Transform
Analyze
Graphs
Utilities
Add-ons
Run
Tools
Window
Help
Active: plan
CONJOINT
Conjoint Analysis in SPSS - You...
IBM SPSS Statistics Processor is ready

```
CONJOINT PLAN='C:\Users\P Venkatesh\Desktop\Analysis\plan_venki_class'
/DATA='C:\Users\P Venkatesh\Desktop\Analysis\data_venki_class'
/SEQUENCE=Card1 TO Card16
/SUBJECT=ID
/FACTORS=Drought (LINEAR MORE)
    Podborer (LINEAR MORE)
   Height (LINEAR MORE)
    Duration (LINEAR Less)
Yield (LINEAR MORE)
/PRINT=SUMMARYONLY.
```

Suggested readings

Lee P Y, Lusk K, Mirosa M, and Oey I (2015). An attribute prioritization-based segmentation of the Chinese consumer market for fruit juice. Food Quality and Preference, 46: 1–8.

IBM SPSS Conjoint 22 Available at: http://www.sussex.ac.uk/its/pdfs/SPSS_Conjoint_22.pdf.

Louviere, J J. (1991), "Analyzing Decision Making: Metric Conjoint Analysis", Sage University Paper Series on Quantitative Applications in Social Sciences, Series No. 07-067. Newbury Park, California.

Annunziata A and Vecchio R (2013) Consumer perception of functional foods: A conjoint analysis with probiotics. Food Quality and Preference, 28(1): 348-355

Praveen KV, Kuar S, Singh D R, Arya P, Chaudhary K and Kumar A (2013) A study on economic behaviour, perception and attitude of households towards traditional and modern food retailing formats in Kochi. Indian Journal of Agricultural marketing ,27 (2) 142-151

Fundamentals and Applications of Choice Experimental Methods

Yashodha
Interantional Water Management Institutte, New Delhi

The methods of valuation of non-marketed goods have become crucial when determining the costs and benefits of public projects. Revealed preference and stated preference methods are two main branches in the non-market good evaluation. Revealed preference methods infer the respondent value of the good by studying actual/revealed behavior. Two well-known methods are hedonic pricing and travel cost methods. Generally, these methods use the actual choices of respondents in evaluating the value of the non-market value. However, evaluated value under the revealed preference method doesn't include non-use value of the good such as existence value, altruistic value, bequest value. Particularly when dealing with the product which is improved with attributes that don't exist in the present state of the good, however, we want to know the value for such attributes such that it directs the policy investment accordingly. Therefore, stated methods become an attractive and growing interest in these methods.

On the other hand, stated preference methods assess the value of non-market goods by using individuals' stated behavior in a hypothetical setting. Stated methods include many different approaches such as conjoint analysis, contingent valuation method (CVM), and choice experiments. In this lecture, we focus on the Choice experiments (CE). It is very popularly used for valuing non-market goods such as ranging from health and environmental applications to transport and public infrastructure projects. The value derived from CE results is used in both cost-benefit analyses and litigations related to damage assessments. This chapter aims to provide the fundamentals of choice design and its applications in research developments to the valuation of non-market goods. We aim to cover, the basic concept of choice experimentation, choice designs, implementation, econometric analysis of CE data.

Rational behind CE

In a choice experiment, individuals are given a number of alternatives and asked to choose their preferred alternative among several alternatives. This sequence is repeated over with several different combinations of alternatives. Each alternative is described by a number of attributes and generally, one of the attributes is a monetary attribute. The way alternatives were placed and described such that when individuals make choices, they implicitly make trade-offs between the levels of the attributes in the different alternatives presented to them. From these traded choices we can derive attribute preference which carries intrinsic value which could be, we can estimate the value for these attributes. Box 1 represents one example of alternative presents in a choice set.

An economic theory which guides the individual choice behavior considers that individual derive the utility from the characteristics of the good. An example of a tourist decides to visit a national park among different parks based on the amenities (characteristics) it offers, like availability of tents, distance, staying facilities. The choice decision of individuals is discrete in nature to choose a park based on park amenities.

Box 1

This is an example choice set presented to farmers in Karnataka to elicit their preference for different wildlife friendly cultivation program (WLFP). This includes 2 alternatives WLFP and a status quo with 3 attribute namely, Land, years.

	WLFP 1	**WLFP 2**	**Status Quo**
Proportion of land	25%	75%	Present cultivation status
Years committed	4years	8years	

Therefore, the individual chose an alternative j over other alternatives i that maximized their utility. Expressing this in the indirect utility function,

$$V_j\left(A_j, y - p_j\; C_j\right) > V_i(A_i, y - p_i\; C_i), \qquad \forall\; i \neq j \qquad (1)$$

Where C represents the alternatives options, p price related to alternative and A represents the fixed preference towards the alternatives. Equation (1) expresses a choice behavioral in a deterministic way. However, the individual choices are also driven by some randomness which re beyond the characteristics/levels.

Random utility frame (RUF) work considers that individual utility has two components, 1) a deterministic component – where utility is a function of goods attributes, 2) stochastic component - where the choices are influences by some randomness which is beyond the good characteristics. Under RUF, the individual's choices are modeled as,

$$V_j\left(A_j, y - p_j\; C_j, \varepsilon_j\right) > V_i(A_i, y - p_i\; C_i, \varepsilon_i) \qquad (2)$$

Where, ε_i is a random probability distribution,

In terms of probabilities, we can express it as,

$$P\left\{\partial_j = 1\right\} = P\{\, V_j\left(A_j, y - p_j\; C_j, \varepsilon_j\right) > V_i(A_i, y - p_i\; C_i, \varepsilon_i)\} \qquad (3)$$

The exact specification of the econometric model depends on how the random elements, ε, enter the conditional indirect utility function and the distributional assumption.

Advantage of CE

Compare to other valuation methods, the CE has gained popularity due to the following potential advantage.

Similar to contingent valuation techniques, CE also elicits the preference from the trade of setting to capture the passive values, however, CE uses experimental design theory that increases the statistical efficiency of the parameters estimated.

The choice format that the respondents face in CE are often similar to those consumers face in markets. This makes it relatively easy for the respondents and attributes can be customized such that they are realistic for respondents and the context undervaluation.

CE offers flexibility to estimate the value for changes in a single character as well as values for multiple changes in characteristics. Therefore, we can have a response for each change in the characteristics rather than a single value and thus offer more information about respondents' behavior.

Because characteristics are experimentally manipulated and presented to respondents, they are typically exogenous, not collinear, and can reflect characteristic levels outside the range of the current market or environment. This is in contrast with revealed preference data that are often collinear, may be limited in variation, and maybe endogenous in explaining choices (Holmes et al 2017).

Since CE uses a possible combination of attributes, it offers the possibility of testing for internal consistency and also avoids several biases such as anchoring bias.

Designing a choice experiment

Before even entering to design the choice experiment, one should devote their time to think whether the CE is a suitable method to answer the question at hand (Alpízar et al, 2011). Once decided to go for CE, the first step is to analyze the dimension of the problem at hand and identify the appropriate attributes and potential changes that can be relevant at for the geography of the good. While identifying, it should be thought who gets affects with the changes in these attributes like, farmers or consumers or policymakers. Identifying the attributes and levels is an iterative process. It requires, structural discussion with the experts, review the similar literature, and conduct focus group discussion with the population where these changes are values to check the relevancy of attributes as well as levels. This is a crucial step, one should spend significant time and effort in scoping the problem and setting the appropriate attributes and levels.

In the second step, once attributes and levels have been determined, the researcher must determine the number of alternatives to present in each choice set and how many numbers of choices set to be presented. It is generally recommended to keep the status quo as one of the alternatives so that it allows us to estimate utility functions that represent changes from baseline conditions. The other alternatives are changes from the status quo and a number of alternatives to be added the choice set depends on the number of alternatives presents in the real world for the specific good at hand. The number of choices set to be asked per individual depends on degrees of freedom(df) require to identify the model, nevertheless, df requirement is usually satisfied in most cases. In most cases, number of choice sets depends on the complexity of the attributes and judgment of the researcher based on the pilot.

The third step is designing a choice set. Given the selected attributes and their levels, how to allocate attribute levels to alternatives create choice sets. This is a technical step, there are several designs available in the CE literature, however, the choice of what parameters to be estimated and whether or not we have prior information in these parameters. An important thing to keep in mind is sufficient independent variation in attribute levels within as well as across alternatives. That is, to minimize the correlation between attribute levels across the alternatives such that it allows us to identify the parameters. A good choice design must balance the attribute level within and across the alternatives such that it is statistically efficient and minimize the standard error of the parameters. Well know choice designs and a short description are presented in the following table

Select the statistically efficient design (mostly lower D-error), and it is generally recommended to use the non-zero prior designs, as they reduce the number of respondents to achieve the specific precision.

Design	**Advantage**	**Limitation**
Orthogonal Designs	Linear in parameter	
Orthogonal Full Factorial	Combines every level of each attribute with every level of all other attributes Orthogonal for both main and interaction effect	The exponential increase in the number of alternatives and choice set with an increase in attributes and attribute levels.
Orthogonal Fractional Factorial Designs,	Select subsets of attribute combinations from full factorial design order interaction terms	In reducing the design size, it omits many interaction effects
Efficient design	Non-linear parameter	
Optimal Orthogonal Designs	Attributes within alternatives are orthogonal and minimal overlap between alternatives that can avoids dominated alternatives	All preference parameters are set to zero
Nonzero Priors Designs	Preference parameters were incorporated in the design that increases the efficiency of the design	If incorrect priors were used, the selected design might not be most efficient

Implementation

After the choice design is set out, choice cards are to be prepared. Since respondents have to make a trade off across different attributes and their levels, the choice cards need to be prepared pragmatically. For better response, most choice cards are designed pictorially and discretely for easy identification of the attributes and levels. it is important to make choice cards.

Before the choice card presentation, it is recommended to develop a small induction about the aim of the experiment, a description of attributes and levels. The choice cards could be administered very meticulously, not to mistake/repeat the same cards over and over. Given the evidence that choice experiments are prone to strategic behavior or hypothetical bias. Many research evidence should that 'cheap talk' and honest priming helps to mitigate strategic behavior.

Estimation methods

After the CE data collection, the preference parameters are estimated using different econometric models. As mentioned before, these models vary depending on the distribution assumed on the error component.

Multinomial logit model (MNL)

MNL models assume that the errors are independently and identically distributed following a Type 1 extreme value (Gumbel) distribution. MNL is the simplest model at the same time very restrictive. That is, it assumes that the alternatives are independent (independent irrelevant alternatives) and respondents have similar tastes and preferences (no preference heterogeneity). These are very strong assumptions.

Latent class model

In this model, it is assumed that respondents belong to different preference classes that are defined by a small number of segments (s). The choice heterogeneity is assumed to exist between these segments, but within this segment, the preference parameters are assumed to be constant.

LCM approach provides information on factors that affect or result in preference differences. That is, the parameters in the segment membership function indicate how the probability of being in a specific segment is affected by age, wealth, or other elements included in the segment membership function. One issue with latent class models is the choice of a number of classes. The determination of the number of classes is not part of the maximization problem and it is an iterative process.

Random parameter model

This is an advanced model in identifying and modeling the respondent's heterogeneity. This captures the individual preference heterogeneity, where it assumes that the taste and preferences of each individual vary. Therefore, estimated parameters vary from one respondent to another. These are complex models, however, with the advancement in computational ability of the computer, several algorithms exist to estimates these models.

Content Analysis, Thematic Analysis and Hands-on session with NVIVO

P Sethuraman Sivakumar
ICAR – Central Tuber Crops Research Institute, Thiruvananthapuram

Qualitative approaches are increasingly popular among social scientists in agriculture, due to their realistic and constructivist approach in analysing complex social phenomenon. Thematic analysis is one of the popular qualitative approaches, used for identifying, analysing, and reporting patterns within data concerning a social phenomenon (Barun and Clarke, 2006). Though thematic analysis is considered as a foundational method for qualitative analysis, many social researchers used it to assist other forms of qualitative analysis (Holloway and Todres, 2003), primarily due to inadequate methodological rigor and lack of clarity in implementing this methodology (Nowell et al., 2017). The purpose of this paper is to introduce thematic analysis as a research method for assessing large volumes of narrative data to derive meaningful interpretation about the phenomenon under study.

Content Analysis Vs Thematic Analysis

In general, the qualitative approaches seek to understand a phenomenon from the people who are experiencing it, through a structured research framework. Along this framework, there is a considerable overlap among these approaches in terms of methods, procedures, and techniques. Both qualitative content analysis and thematic analysis share same goal of assessing the phenomenon by breaking the narrative text into relatively small units of content and submitting them to descriptive treatment (Sparker, 2005). However, there are few functional differences among these methods (Vaismoradi et al., 2013) (Table 1).

Table 1. Differences between content analysis and thematic analysis

	Content analysis	Thematic analysis
Purpose	To describe the characteristics of the narrative of a specific phenomenon by examining who says what, to whom, and with what effect.	To identify common threads or themes from that extend across an entire narrative
Research approach	Mixed methods – Both qualitative and quantitative	Qualitative
Focus of research design	Description and more interpretation than thematic analysis	Minimal description and interpretation
Consideration of context of data	Danger of missing context- only the frequency of codes is	Combines analysis of the

	counted to find significant meanings in the text	meaning derived from data within particular context
Content type	Considers either manifest (developing categories) or latent contents (developing themes).	Considers both manifest and latent contents
Nature of theme	Derived based on frequency of occurrence of content; represents only surface meaning	Abstract, mostly latent and derived through an intense qualitative process
Mapping of themes	No	Yes
Assessment of reliability of coders	Assessed through inter-coder reliability	Code book, Audit Trails

(Adapted from Vaismoradi et al., 2013)

When to use Thematic Analysis

Thematic analysis is suitable for understanding a phenomenon through stakeholder's views, opinions, knowledge, experiences or values, derived from a set of qualitative data. The common sources of qualitative data are audio/video recorded personal interviews; audio-recorded telephonic interviews; blogs on a specific topic; social media posts on a specific topic; case studies/success stories; newspaper reports; other audio or video recordings of events, views, etc.; reports; policy documents; and feedback forms.

Different approaches to thematic analysis

Themes or patterns within data can be identified in one of two primary ways in thematic analysis: in an inductive or 'bottom up' way or in a theoretical or deductive or 'top down' way (Braun and Clarke, 2006).

Inductive approach – Data- driven analysis which involves coding the data without trying to fit it into a pre-existing coding frame, or the researcher's analytic preconceptions

Deductive approach – Analyst-driven approach which follows researcher's theoretical or analytic interest in the area.

Besides, thematic analysis also looks into nature of themes – semantic and latent (Braun & Clarke, 2006).

Semantic themes – Focus on the surface meanings of the data confined to what a participant has said or what has been written.

Latent themes – Analysis extend beyond respondents views to identify or examine the underlying ideas, assumptions, and conceptualisations and ideologies, which shape semantic content of the data

Advantages of thematic analysis

Thematic analysis has several advantages over other qualitative methods (Braun and Clarke, 2006; King, 2004; Nowell et al., 2017).

Highly flexible approach customised to the needs of researchers to assess complex qualitative data.

Useful method for examining the phenomenon from perspectives of different research participants, which help in highlighting similarities and differences, and generating unanticipated insights.

Useful for summarizing key features of a large data set.

Disadvantages

Though thematic analysis has several advantages, it has few limitations too (Braun and Clarke, 2006; Holloway and Todres, 2003; Nowell et al., 2017).

The lack of substantial literature on thematic analysis compared to other qualitative methods like grounded theory, ethnography, and phenomenology.

It does not allow researcher to make claims about language use

Inconsistency and lack of coherence when developing themes derived from the research data

Steps in conducting Thematic Analysis

Considering the strengths and limitation of thematic analysis as a qualitative approach, a systematic approach and trustworthy approach is proposed following the guidelines suggested by Braun and Clarke (2006) and Nowell et al., (2017) on the trustworthiness criteria suggested by Lincoln and Guba (1985).

Step 1: Define the research problem and question

The researcher develop the research problem in a systematic way by specifying appropriate research goals, along with clear, concise and sound research questions.

Step 2: Sampling strategy

The sampling strategy is the plan devised by the researcher to ensure that the sample chosen for the research work represents the selected population. Robinson (2014) proposed a four-point sampling process for systematically selecting adequate samples for obtaining quality results. It involves (i) Defining a sample universe – inclusion/ exclusion criteria for respondents; (ii) Selecting adequate samples; (iii) Choosing relevant sampling method and (iv) ways of sourcing samples.

Step 3: Collect data in a systematic way

The researcher collects the data in a systematic way following the sampling strategy in an unbiased and error free manner. After collecting the data, a data corpus containing all the information gathered for the thematic analysis is prepared.

Step 4: Transcription and translation of verbal data

The collected data is often in the form of audio or video forms, which need to be converted into textual form for analysis. It involves two processes – Transcription and translation.

Transcription is the process of converting the verbal data i.e., interviews, audio/video clips and speeches into written form of the same language (Barun and Clarke, 2006) while translation is the process of translating the transcribed verbatim into English or any other language in written form. A systematic process of objective ways of doing transcription and translation is described by Chen and Boore (2009).

5. Familiarising with data

At this stage, the researcher makes a quick glance at the verbatim/ transcripts as a whole and takes notes from first impressions. Then, the verbatim/transcripts are thoroughly read line by line by looking for meanings and patterns from the data. The trustworthiness of data can be ensured through (i) Triangulating different data collection modes; (ii) documenting theoretical and reflective thoughts, (iii) documenting thoughts about potential codes/themes and (iv) Storing data properly along with all records including field notes, transcripts, and reflexive journals (Nowell et al., 2017)

6. Generating initial codes or labels

After initial reading, identify the data extracts or specific word(s) which represent a dimension of the research problem. The coding process is performed systematically across the entire data set by collating data relevant to each code.

During the coding process, the researcher highlights or underlines the specific data extracts while reading the transcript. The researcher can also write them down in a separate notebook for grouping in the later stages.After identifying the data extract, assign a code to it based on your perception of what it signifies.Group all the codes along with relevant data extracts and prepare a long list of codes and data extracts.

At this phase, trustworthiness can be ensured through (i) Peer debriefing; (ii) Researcher triangulation; (iii) Reflexive journaling; (iv) Use of a coding framework; (iv) Audit trail of code generation; and (v) Documentation of all team meeting and peer debriefings (Nowell et al., 2017).

7. Searching for themes

This phase involves sorting the identified codes into preliminary levels of themes based on their perceived closeness, and pooling all the relevant coded data extracts into the identified themes. A theme is essentially a coherent and meaningful pattern in the verbatim/ transcript relevant to the research question. It is technically a construct or a dimension of the construct. Visual techniques like mind maps, tables and cards may be used to pool the relevant codes into preliminary themes.

At this phase, trustworthiness can be ensured through (i) Researcher triangulation and (ii) Storing detailed notes about development and hierarchies of concepts and themes (Nowell et al., 2017)

At the end of this phase, the preliminary themes, and sub-themes, and all extracts of data that have been coded in relation to them are identified and plotted.

8. Reviewing themes

At this stage, the researcher checks if the themes work in relation to the coded extracts and the entire data set, generating a thematic map. The reviewing and refining are performed following Patton's (1990) dual criteria for judging categories - internal homogeneity and external heterogeneity

Level 1. Internal Homogenity - Reviewing at the level of the coded data (Reviewing the codes and data extracts)

Read all the collated codes and respective data extracts for each theme and sub-theme to check if the data forms a coherent pattern.

If the main and sub- themes do not fit, you would rework your theme, creating a new theme, finding a home for those extracts that do not.

Level 2. External Homogenity - Over all reviewing of themes with the data set Consider each theme in relation to your data corpus.

Generating and checking Thematic map - Do the relationships between the themes reflect the meaning of your data as a whole?

At this phase, trustworthiness can be ensured through (i) Researcher triangulation; (ii) vetting of themes and subthemes by team members; (iii) conducting test for referential adequacy by returning to raw data (Nowell et al., 2017)

Step 9: Defining and Naming Themes

This phase captures the essence of what each theme is about and what aspect of the data each theme captures. In this phase, the themes, sub-themes are examined carefully to see that they are coherent and internally consistent. Each theme get a name - concise, punchy and immediately give the reader a sense of what the theme is about. The final thematic map is drawn with description.

At this phase, trustworthiness can be ensured through (i) researcher triangulation; (ii) peer debriefing; (iii) team consensus on themes; (iv) documentation of team meetings regarding themes; and (v) Documentation of theme naming (Nowell et al., 2017).

Step 10: Producing the Report

The final phase begins once the researcher has fully established the themes and is ready to begin the final analysis and write-up of the report (Braun and Clarke, 2006). The write-up of a thematic analysis should provide a concise, coherent, logical, non-repetitive, and interesting account of the data within and across themes (Braun and Clarke, 2006).

At this phase, trustworthiness can be ensured through (i) member checking, (ii) peer debriefing; (iii) describing process of coding and analysis in sufficient details; (iv) thick descriptions of context; (v) description of the audit trail; (vi) report on reasons for theoretical, methodological, and analytical choices throughout the entire study (Nowell et al., 2017).

Software for thematic analysis

Thematic analysis is often performed manually since it involves identification of semantic and latent themes. With the recent advances in Natural Language Processing, Verbatim analysis or text analytics and Word embedding applications, many qualitative analysis software have inbuilt capacities to do thematic analysis. Popular software used for thematic analysis are Nvivo (https://www.qsrinternational.com/nvivo-qualitative-data-analysis-software/home), QDS Miner (https://provalisresearch.com/products/qualitative-data-analysis-software/freeware/) and ATLAS.ti (https://atlasti.com/). The QDA Miner Lite- free version of QDA Miner is popular freeware for conducting thematic analysis.

References

Braun, V., and Clarke, V. (2006). Using thematic analysis in psychology. Qualitative Research in Psychology, 3: 77–101.

Chen, H-Y. and Boore, J.R.P. (2009.) Translation and back-translation in qualitative nursing research: Methodological review. Journal of Clinical Nursing 19:234–239

Holloway, I., and Todres, L. (2003). The status of method: Flexibility, consistency and coherence. Qualitative Research, 3: 345–357.

King, N. (2004). Using templates in the thematic analysis of text. In C. Cassell & G. Symon (Eds.), Essential guide to qualitative methods in organizational research (pp. 257–270). London, UK: Sage.

Lincoln, Y., and Guba, E. G. (1985). Naturalistic inquiry. Newbury Park, CA: Sage.

Nowell, L. S., Norris, J. M., White, D. E., and Moules, N. J. (2017). Thematic Analysis: Striving to meet the trustworthiness criteria. International Journal of Qualitative Methods, 16 (1), 1-13.

Robinson, O.C. (2014.) Sampling in interview-based qualitative research: A theoretical and practical guide. Qualitative Research in Psychology, 11:25–41.

Vaismoradi, M., Turunen, H., and Bondas, T. (2013). Content analysis and thematic analysis: Implications for conducting a qualitative descriptive study. Nursing and Health Sciences, 15(3), 398–405.

Social Network Analysis- Theory and Practice

Sreeram Vishnu
Regional Agricultural Research Station, Kerala Agrcultural University, Wayanad

Social network analysis (SNA) is an effective tool in understanding the social relations and interactions of the individuals in the group (Bougatti et al 2009). It helps to understand the actors and the relationship between them in a specific social context (Clark 2006). Initially it was used in field of sociology, psychology and anthropology. With the advancement in graph theory (mathematics) and computing knowledge, SNA softwares are available to map and quantify networks.This has been utilized in various dimensions in social sciences. SNA studies also differs from conventional survey-based studies in the definitions of boundaries, samples and populations.

Evolution of SNA

The first true formulations of social network analysis, in which the metaphor was taken seriously as the basis for developing a range of sociological concepts, took place in the American social psychology in the 1930s (Scott,1988). In his book, The Development of Social Network Analysis: A Study in the Sociology of Science origin (2004), Freeman argued that it was three independent researchers from various fields, whose works laid the foundation for the field SNA. The first among them was a psychiatrist J L Moreno and a psychologist Helen Jennings who were credited for the technique of sociometry. Their findings were based mainly on the studies, first among the inmates of a prison (Moreno et al.,1932) and later among the residents in a reform school for girls (Moreno, 1934). Secondly it was the works of an anthropologist, W. Lloyd Warner who designed the little known "bank wiring room" study, a social network component of the famous Hawthorne studies on industrial productivity (Roethlisberger and Dixon, 1939). The third one was the German psychologist, Kurt Lewin who developed a structural perspective and conducted social network research in the field of social psychology (Lewin and Lippit, 1938). Lewin (1951) further employed mathematical techniques such as topology and set theory to explore the social space, but Kőnig's (1936) graph theory provided the crucial breakthrough in application of mathematical concept in sociometric analysis. According to Freeman (2004), by 1970 the social network analysis gained traction among social scientists. The following figure (Figure 1) depicts the lineage of Social Network Analysis. Social network analysis in the current form is an amalgamation of socio-metric technique and graph theory.

It could be seen that initially SNA was mainly used in fields of sociology, psychology and anthropology. With the advancement in graph theory (mathematics) and computing knowledge, SNA tools have been developed to map and quantify networks. Some of the important terms in widely cited in the SNA literature are detailed in the following table (Table 1). These basic terminologies used in SNA were basically derived from the graph theory. Each of them have distinct interpretations while being quantified.

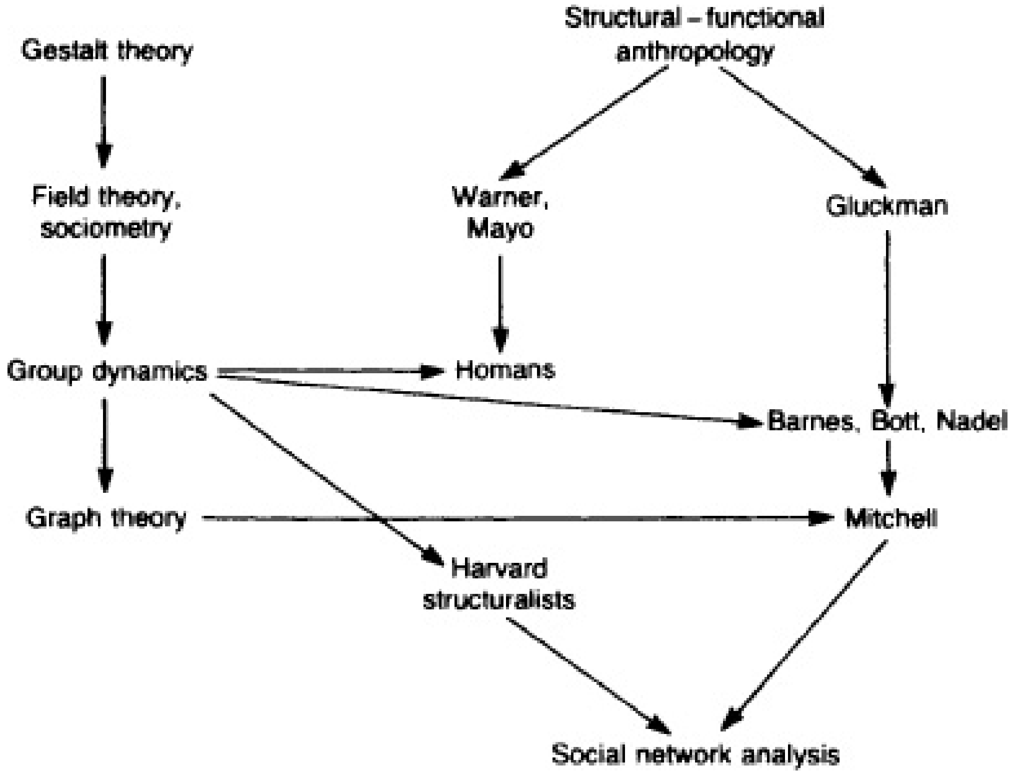


Fig 1: Lineage of Social Network Analysis

Table 1: Elements of Social Network Analysis

Sl. no.	Element	Definition
1	Node	Any individual, organization, or other entity of interest
2	Ego	Actor of interest within a network
3	Geodesic distance	Shortest path connecting two nodes
4	Network	Graphical representation of relationships that displays points to represent nodes and lines to represent ties; also referred to as a graph
5	Ego Network	Network that only shows direct ties to the ego and not between alters
6	Network size	Total number of nodes in a network
7	Walk	A series of interconnected nodes
8	Path	Walk where each node and line is only used once
9	Network centralization	Degree to which a network revolves around a single node
10	Network density	Nodes that are actually tied as a proportion of all possible ties in a network
11	Centrality	Measure of the number of ties that a node has relative to the total number of ties existing in the network as a whole; centrality measures include degree, closeness, and betweenness

12	Degree	Number of ties a node has to other nodes
13	Closeness	Measure of reciprocal of the geodesic distance (the shortest path connecting two nodes) of node to all other nodes in the network
14	Betweenness	Number of times a node occurs along a geodesic path
15	Network boundary	Natural delineation between actors and relationships, or artificial limit set by a researcher

Source: Borgatti (1997, 1998), Davies (2004), Hanneman and Riddle (2005)

A social network is a set of actors (or nodes) that may have relationships (or ties) with one another (Hanneman, 2001). SNA is based on a so-called adjacency matrix or connection matrix (Bender et al., 2015). Uni-modal networks (considering only a single attribute at a point such as information flow among the actors) were constructed out of the data, by compiling it in a square (n x n) matrix of n actors (nodes). A simple non directional tie (for example mutual linkage) between two nodes is represented as nij = nji = 1 in the matrix. A directional tie, for example a flow of information from node i to j but not from j to i is represented as nij = 1 but nji = 0. Directed ties in a network graph are indicated by arrows, and an undirected graph shows only the lines between nodes. Data for SNA are commonly based on measurements of relationships between actors and sets of actors, in addition to the attributes of individual actors. Some of the important network parameters are discussed below.

Centrality is the measure of how close an individual is to the centre of the action in a network (Chan and Liebowitz, 2006). It refers to locations of positions or points in a network and is applied to the overall structure of the network as a whole (Freeman, 1979). There are three important measures of centrality which are widely reported by various researchers. viz., degree, betweenness and closeness.

Degree centrality (Cd) measures the number of ties that a node has relative to the total number of ties existing in the network as a whole, or

$$C(dni)= \beta i \ (ni)/ (N-1)$$

Where ni denotes ith node in the network and βi (ni) denotes the number of ties to ni and (N – 1) represents the size of the network less the node of interest. The degree of a node is viewed as important as an index of its potential ability for direct interaction. Again it can be in degree or out degree depending on the number of choices one node receive or make respectively. The former is an indicator of prominence of a node in a network while the latter is suggestive of influence depending on the nature of information receiver in the network.

Betweenness centrality (Cb) of an actor or node is a measure of how often an actor lies on the shortest path between two other actors or nodes. It is calculated as,

$$Cb\ (i) = \Sigma\ j<k\ g\ jk\ (i)/gjk$$

Where gjk = the number of geodesics connecting j and k, and gjk(i)= the number of geodesics that actor i is on. This measure is based on the frequency with which a node falls between pairs of other nodes on the geodesic paths connecting between them. A node occupying such a strategic position can influence the group by withholding or distorting the information in transmission. They have the responsibility of maintenance of the communication or can function as potential group co-ordinators (Cohn and Marriott, 1958).

Closeness centrality (Cc) measures the reciprocal of the geodesic distance (the shortest path connecting two nodes) of node ni to all other nodes in the network, or

Cc (ni)-1= $\sum_{i=1}^{N} d(ni,nj)$

where, d(ni,nj) denotes the number of ties in the geodesic paths linking ni and nj. An actor with high closeness centrality will be closely connected to many actors, and thus be in a position to receive information or other resources from the network (Spielman and Kelemework, 2007).

Further the properties of whole network are captured by deriving unique parameters like density, average degree and number of ties. Density of a network (D) indicates the degree to which members are connected to all other members. It is calculated as the ratio of the number of actual links in a population to the number of possible links in the population.

D=µ/ N(N-1)/2

Where µ denotes the total number of lines (ties) present and N is the number of nodes in the network. Information in the low-density graph can flow through only one route, whereas information in the high-density graph can flow from and to a number of different actors (Haythornthwaite, 1996). Total number of ties gives an idea on the number of linkages present which intuitively explains the density of the network.

Application of SNA in Social sciences

SNA has been used to map and quantify the value chains (Lazzarini, 2001; Borgatti and Li, 2009; Trienekens, 2011; Bellamy and Basole, 2012) , monitoring and impact assessment (Ekboiret al., 2011), and to understand adoption of technology (Matsuskhe, 2008; Magnan et al., 2015) and rural innovations (Spielman et al., 2010). Bartholomay & Chazdon, (2011) employed SNA to examine the structure and dimensions of extension outreach programme. SNA allows for the study of relationships among multiple and diverse actors in a system (Borgatti, 2006) and their influence on generation, exchange, and use of information and knowledge (Yauney et al., 2012, Cross et al., 2003). Spielman et al., (2010) utilized social network analysis for mapping the innovation network. The study utilized network map and centrality measures for illustrating and identifying the key actors.With the set of analytical tools in SNA, the networks of relationships could be mapped and provides an important means of assessing and promoting collaboration in strategically important groups (Cross et al., 2003). Thus the network analysis could be visualized by network maps (Freeman, 2000; Liebowitz, 2005) and quantitatively measured by degree and other centrality measures (Freeman, 1979; Landherr et al., 2010).

Software packages available for SNA analysis

A number of software packages available for SNA analysis. In general network analysis software can be classified into two types; packages based on graphical user interfaces (GUIs) and packages for scripting or programming languages. GUIs are easier to learn and execute while scripting packages are powerful. The most widely used GUI packages are UciNet, Pajek, Gephi, muxViz, NetMiner, GUESS, ORA. Netminer (Phython), igraph (package for R and Python) are couple of scripting based packages. Netdraw is an open source and widely used network mapping tool. Both free and commercial versions of different software are available. Though open source packages are difficult to learn they have much wider functionality and features than the commercial ones and lot of training, tutorials and support groups are available for them. The software mentioned above could be used for visualizing networks through network maps and quantitatively measure network parameters.

References

Bartholomay, T., & Chazdon, S. (2011). December 2011 Article Number 6FEA9 Mapping Extension ' s Networks : Using Social Network Analysis to Explore Extension ' s Outreach, 49(6), 1–14.

Bellamy, M. ., & Basole, R. . (2011). Network Analysis of Supply Chain Systems: A Systematic Review and Future Research. Systems Engineering, 14(3), 305–326.

Bender, M,E., Keil, T., Bender, M,E., Molyneux, D. (2015).Using Co-authorship Networks to Map and Analyse Global Neglected Tropical Disease Research with an Affiliation to Germany. PLoS Negl Trop Dis.; 9: e0004182.

Borgatti, S., & Li, X. (2009). On Social Network Analysis in a Supply Chain Context. Journal of Supply Chain Management. 45(2): 1–17.

Borgatti. S.P. 2006. Identifying sets of keyplayers in a social network. Computational and Mathematical Organization Theory 12(1):21–34

Borgatti, S. (1998). Social network analysis instructional web site. Retrieved from http://www.analytictech.com/networks/. (Accessed 01/03/2015)

Borgatti, S. (1997). Structural holes: Unpacking Burt's redundancy measures. Connections, 20(1): 35–38.

Chan, K and Liebowotz, (2006). The synergy of social network analysis and knowledge mapping: a case study. International Journal of Management and Decision Making, 7(1): 19-35.

Clark, L (.2006).Building farmers' capacities for networking (Part II).Strengthening agricultural supply chains in Bolivia using network analysis. Knowledge Management for Development Journal,2(2):19-32.

Cohn, B,S., and Marriott, M.(1958). "Networks and centres of integration in Indian civilization." Journal of Social Research I: 1-9.

Cross, R., Parker, A. and Sasson, L. (Eds.) (2003) Networks in the Knowledge Economy,Oxford University Press, New York, NY, pp.82–105.

Davies, R. 2004. Scale, complexity and the representation of theories of change: Part II. Evaluation, 11(2): 133–149.

Ekboir, J., Canto, G. B., & Sette, C. (2011). Monitoring the composition and evolution of the research networks of the CGIAR Research Program on Roots, Tubers and Bananas (RTB). Available online on http://www.cgiar-ilac.org/files/ilac_report_research_networks_rtb_0.pdf. Accessed on 11-07-2015.

Freeman, L. (2004). The development of social network analysis. A Study in the Sociology of Science.

Freeman, L. C. (2000). Visualizing social networks. Journal of social structure,1(1): 4.

Freeman, L. C. (1979). Centrality in Social Networks Conceptual Clarification, Social Networks. 1: 215–239.

Hanneman, R. 2001. Introduction to Social Network Methods, Retrieved from http://www.faculty.ucr.edu/~hanneman/. (Accessed on 8/03/2015)

Hanneman, R.A. and Riddle, M. (2005). Introduction to social network methods. University of California, Riverside.

Haythornthwaite, C. (1996).Social network analysis: An approach and technique for the study of information exchange. Paper presented at the 1996 ALISE conference, San Antonio, Texas.

Kőnig, D. (1936). "Gráfokésmátrixok", MatematikaiésFizikai Lapok, 38: 116–119.

Landherr, A., Friedl, B., & Heidemann, J. (2010). A Critical Review of Centrality Measures in Social Networks. Business & Information Systems Engineering, 2(6), 371–385. doi:10.1007/s12599-010-0127-3

Liebowitz, J. (2005). Linking social network analysis with the analytic hierarchy process for knowledge mapping in organizations. Journal of Knowledge Management, 9(1), 76–86. 4

Lazzarini, S., Chaddad, F., & Cook, M. (2001). Integrating supply chain and network analyses: the study of netchains. Journal on chain and network science, 1(1), 7-22.

Lewin, K. (1951). Field theory in social science. Retrieved from http://agris.fao.org/agrissearch/search.do?recordID=US201300602463. (Accessed on 01/03/2015)

Lewin, K., and Lippitt, R. (1938). An experimental approach to the study of autocracy and democracy: A preliminary note. Sociometry, 1(4): 292-300.

Magnan, N., Spielman, D. J., & Lybbert, T. J. (2015). Information Networks among Women and Men and the Demand for an Agricultural Technology in India. IFPRI Discussion paper 01411.

Matuschke, I. (2008). Evaluating the impact of social networks in rural innovation systems: An overview. IFPRI Discussion Paper, (November), 26.

Moreno, J. L., Whitin, E. S., and Jennings, H. H. (1932). Application of the group method to classification. National committee on prisons and prison labor.

Moreno, J. L. (1934). Who shall survive (Vol. 58). Washington.

Roethlisberger, F. J., and Dickson, W. J. (1939). Management and the Worker. Psychology Press.

Scott,J.(1988).Social network analysis and intercorporate relations. Hitotsubashi Journal of Commerce and Management,23(1): 53-68.

Spielman. D. J. and Kelemework, D. (2007). Measuring agricultural innovation system properties and performance: illustrations from Ethiopia and Vietnam. IFPRI discussion paper No. 851. Washington, DC: IFPRI.

Spielman, D. J., Davis, K., Negash, M., & Ayele, G. (2010). Rural innovation systems and networks: findings from a study of Ethiopian smallholders. Agriculture and Human Values, 28(2), 195–212. http://doi.org/10.1007/s10460-010-9273-y

Trienekens, J. H. (2011). Agricultural value chains in developing countries a framework for analysis. International Food and Agribusiness Management Review, 14(2), 51–82.

Yauney, J.,Thangata, P.,Droppellmann, K. and Mapila, M. (2012).Who talks to whom? An analysis of information flows in Malawi's agricultural research network. International Food Policy Research Institute.

Application of Psychometrics for Behavioural Research

R.N.Padaria
Division of Agricultural Extension, ICAR-Indian Agricultural Research Institute, New Delhi

Psychometrics refers to psychological measurement. One of the important applications of psychometrics has been construction and validation of scales and tests, which have been immensely used in agricultural extension research to measure psychological variables like attitude, achievement motivation, entrepreneurial orientation, risk orientation, etc.

Approaches and methods of Scale Development

Arbitrary approach: Very often when a researcher or an evaluator needs to understand whether there is difference in degree of judgment or response towards any characteristic or trait, scales are used. Sometimes a scale is developed with an arbitrary collection of statements based on heuristics and subjective judgment of the researcher and is administered to respondents for measuring the characteristics in question. Such scales are called arbitrary scales.

Differential scales approach: There are scales in which the items or statements are selected and rank ordered on a continuum by a group of experts. Such scales are known as differential scales. Various methods of judgments for relative ranking of statements based on Thurstone's principles are used like method of paired comparison or equal appearing intervals. Therefore, differentials scales are also referred as Thurstone's type scales. The person's response to the statement fixes his or her position on the continuum.

Item analysis approach: The statements are selected for a scale based upon its discriminatory power. Likert scale falls under this category. Since the total score of an individual is obtained by summation of scores of responses to all the statements of a scale, such scales are also called summated scales.

Cumulative scale approach: Cumulative scale or Guttman scale lay emphasis upon unidimensionality of a scale and it is checked through scalogram analysis. A scale is unidimensional if the statements of scale fall along a single dimension. There are two techniques for conducting scalogram analysis i.e. Cornell technique and Goodenough technique. A perfect scale makes it possible to reproduce the responses to the individual statements from knowledge of total scores. Scalogram analysis provides an estimate of coefficient of reproducibility, which indicates the percent accuracy with which responses to the various statements can be reproduced from the total scores.

Factor analysis approach: Factor analysis is another approach for scale development. Scale developed through factor analysis is called factor scale. Factor analysis helps to determine the number of latent variables underlying a set of statements. Semantic Differential (S.D.) and the multidimensional scaling are based upon factor analysis.

Comparative and non-comparative scaling techniques

The scaling techniques can be compared as comparative scales and non-comparative scales. The scaling technique in which there is a direct comparison of stimulus objects with each other is known as comparative scale. Paired comparisons, rank order, constant sum scales, Q-sort and other procedures are comparative scales. On the contrary, the stimulus objects are scaled independently of other objects in the stimulus set, it is known as non-comparative scale. It can be continuous rating or itemized rating scales. The itemized rating scale can be classifies as Likert scale, semantic differential scale and staple scale.

Modern approaches to psychological scaling

The psychometric methods could be divided into three major classes viz., psychological scaling; factor analysis, and test theory. Psychological scaling comprises a set of techniques for assignment of quantitative values to objects or events based upon the data obtained through human judgment. Factor analysis methods aim at explaining the observed co-variation among a set of variables. Item Response Theory (IRT) assumes that one or more unobserved (latent) variables underlie the responses to test items in the sense that variation among the individuals on those latent variables explains the observed co-variation among item responses.

Item Response Theory (IRT): Though Classic test theory (CTT) has been the basis for developing psychological scales and test scoring for many decades, Item Response Theory (IRT), is a new approach being applied by psychometricians for explaining and analyzing the relationship between the characteristics of an individual and his/her response to the individual items. Item Response Theory (IRT) emphasizes that besides the item properties like item difficulty and item discrimination, a respondent's response to an item of a psychological scale or test also very much depends upon the standing of the respondent on the psychological characteristic being measured by the item.

IRT provides information about the quality of a scale's item and of the entire scale. There are several important uses of IRT. With item information and test information items having good discriminative ability could be identified. The second important use of IRT is examination of differential item function. It occurs when an item functions differently in different groups of respondents. The third key use is examination of person-fit. Analysis of person-fit identifies people whose response pattern does not fit the expected pattern of responses to a set of items. IRT also facilitates Computerized Adaptive Testing (CAT) intended to produce accurate and efficient assessment of individual's psychological characteristics.

Factor analysis for evaluating dimensionality, internal structure and validity of scale

Factor analysis (FA) is the most important statistical tool for validating the structure of our instruments. There are other components of construct validity that are not addressed by factor analysis. FA is usually a two-stage process. The first stage of FA offers a systematic means of examining inter-relationships among items on a scale. This stage of FA is exploratory factors analysis. Exploratory factor analysis (EFA) is the most common method of evaluating the dimensionality of psychological scales. If all scale items are well correlated with each other at about equal levels, the scale is unidimensional. Exploratory factor analysis (EFA) is useful when a researcher has a few hypotheses about a scale's internal structure. On the contrary, when a researcher has a clear hypothesis about a scale i.e. the number of factors or dimensions underlying its items, links between items and factors, and the association between factors, Confirmatory factor analysis (CFA) is useful. Confirmatory factor analysis (CFA) is a statistical method appropriate for testing whether a theoretical model of relationships is consistent with a given set of data. CFA allows researchers to evaluate the degree to which their measurement hypotheses are consistent with actual data produced. CFA facilitates theory testing, theory comparison, and theory development in a measurement context.

Application of Structural Equation Modeling (SEM) has gained attention for scale development. While EFA is used to study single relationships individually, SEM deals with multiple dependence relationship. Structural modeling refers to the systematic identification of possible relationship among concepts. The structure of relationship is represented with mathematical equations.

Multidimensional scaling (MDS): MDS represents a set of stimuli as points in a multidimensional space in such a way that those points corresponding to similar stimuli are located close together, while those corresponding to dissimilar stimuli are located far apart. The basic idea behind MDS is similarity/dis-similarity data or proximity data obtained by various spatial distance models. Most frequently used model is Euclidean model.. For social science, Non-metric MDS is the most suited one since data are mostly at ordinal level. MDS can be used rigorously in the field of extension in measuring multi-dimensional variables, attitude, perception, semantic differential, positioning of innovation, audience segmentation, targeting, discovering underlying behavioral and personality factors, understanding audience preferences etc. Since it is based on respondents' subjective perception and subjective evaluation it gives us greater understanding of our target clientele and individual differences prevailing among them.

Testing the reliability and validity of the scale

Reliability: It refers to the accuracy or precision of the measuring instruments. Reliability can be defined in terms of relative absence of errors of measurement in a measuring instrument.

Methods to measure reliability: There are different methods of testing reliability of any psychological measurement tools such as Test-retest method, Parallel forms method, and Split- half method.

Test-retest: As the name of the method signifies, the scale to be tested for reliability, is administered to a group of individuals at two points of time (usually at a gap of 15 to 30 days) and the scores obtained are correlated. Value of correlation r gives us the reliability coefficient. Higher the value r, higher is the reliability of the test.

Parallel form or equivalent form: Two separate scales comprising similar items on the psychological object are used simultaneously. Both the tests are equivalent in terms of their items. The two tests on psychological objects are administered one after the other to a group of subjects and the scores on the two scales are then correlated. The value of r obtained is the reliability coefficient and is also known as coefficient of equivalence.

Split- half method- In this method, the test scores on one half of the scale are correlated with the scores on other half. The split- half method is based on the assumption that if the test items are homogeneous and there is internal consistency the scores on any item or set of items on the scale would yield high correlation value with any other item or set of items (the number of items being the same in the two subsets). The r-value worked out in this case is for half of the scale; hence, to have the reliability coefficient of the entire test scale we need to apply 'Spearman-Brown formula'

$$\frac{nr_1}{1+(n-1)r_1}$$

rtt = reliability of the original test

r = reliability coefficient of the subsets

n =number of times the length of the original test-is shortened

Validity: It is the degree to which a measuring instrument measures what it is supposed to measure. Validity refers to the appropriateness of the instruments/test. A test is said to be valid when it measures what it is supposed to measure. Statistically, validity is the problem of common factor variance to total variance. According to Guttman, there are two broad types of validity, i.e., internal and external.

Internal validity expresses a logical relationship between the theoretical and operational definition of the concept under study.

External validity expresses an empirical relationship between the theoretical definition and the operational definition.

Validity has different levels viz, content, criterion, and construct validity. Content validity is evaluated by determining the degree to which the items of a scale/test represent the universe of content of the object phenomenon being measured by it and their adequacy. It is related with the representativeness and adequacy of the context of the test/scale. Criterion validity has two forms i.e. Predictive validity and Concurrent validity. Predictive validity is estimated by showing how accurately we can guess some future performance, on the basis of the measure on external or other criteria, e.g., on the basis of the score on leadership scale of an individual, one can predict accurately his behavior as a manager. While predictive validity is used for forecasting the presence or absence of the trait in future, based on the scores on the criteria obtained today. The concurrent validity is based on simultaneous comparison of scores on one test/scale with other established criteria. Construct validity is the strongest and the most meaningful validity compared to the other two. Construct validity is useful in validating the construct, the theory behind the test. It is evaluated by a determination of the relationship between the test attitude score and other aspects of the individual personality.

Practical Manual

Advances in Research Methodology for Social Sciences

August 31- September 4, 2020

Division of Agriculture Economics, ICAR-IARI, New Delhi

Course Convenor

Alka Singh
Professor and Head
Division of Agricultural Economics
ICAR-Indian Agricultural Research Institute
Pusa Campus, Delhi – 110012
Email: asingh.eco@gmail.com
Phone No. 9871198527

Co-Convenor

R.N. Padaria
Professor
Division of Agricultural Extension
ICAR-Indian Agricultural Research Institute
New Delhi 110 012
E-mail: rabi64@gmail.com

R. R Burman
Principal Scientist
Division of Agricultural Extension
ICAR-Indian Agricultural Research Institute
New Delhi 110 012
E-mail: burman_extn@hotmail.com

Aditya K.S.
Scientist
Division of Agricultural Economics
ICAR-Indian Agricultural Research Institute
New Delhi 110 012
E-mail: adityaag68@gmail.com

Praveen KV
Scientist
Division of Agricultural Economics
ICAR-Indian Agricultural Research Institute
New Delhi 110 012
E-mail: veenkv@gmail.com

Published By

NAHEP- Centre for Advanced Agriculture Science and Technology
ICAR- Indian Agriculture Research Institute, New Delhi
Website: **www.nahep-caast.iari.res.in**