Deutsche Gesellschaft
für Internationale
Zusammenarbeit (GIZ) GmbH

INFINITE OPPORTUNITIES | GERMANY + INDIA

nidm

Training Modules

Databases and Statistics for Disaster Risk Management

Sreeja S. Nair, Anil K. Gupta and Klaus Röder

Training Modules

Databases and Statistics for Disaster Risk Management

Imprint

ISBN: 978–3–944152–11–0

Published by
National Institute of Disaster Management, Ministry of Home Affairs (NIDM)
5–B, IIPA Campus, IP Estate, Mahatma Gandhi Marg
New Delhi 110 002, India
T: +91 11 23702432, 23705583, 23766146
F: +91 11 23702442, 23702446
I: www.nidm.gov.in

and

Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH
Indo–German Environment Partnership
B–5/2 Safdarjung Enclave
New Delhi 110 029, India
T: +91 11 49495353
F: +91 11 49495391
I: www.giz.de

Authors
Sreeja S. Nair
Anil K. Gupta
Klaus Röder

Review
Dr. S.K. Nath, former Director General, Central Statistical Organization, India
Florian Bemmerlein–Lux, Peter O. Bank, ifanos concept & planning, Germany

Acknowledgements to
Dr. Satendra, IFS, Executive Director, NIDM
Dr. Dieter Mutz, Director, IGEP, GIZ

Citation
Nair S.S., A.K. Gupta and K. Röder (2013). Databases and Statistics for Disaster Risk Management.National Institute of Disaster Management and Deutsche Gesellschaft für internationale Zusammenarbeit GmbH (GIZ), New Delhi, 228 p.

Design and Printing
M/s Perfect Impact Designing & Printing Pvt. Ltd., 37/534, New Moti Nagar, New Delhi, September, 2013

Dr. Dieter Mutz
Director
Indo–German Environment Partnership (IGEP) Programme
GIZ–Germany
India Office, New Delhi
September 2013

Message from GIZ–IGEP

Since 2010, GIZ has been collaborating with the National Institute of Disaster Management for implementing the "Environmental Knowledge and Disaster Risk Management" (ekDRM) project, aimed at strengthening capacity building initiatives in knowledge management and risk reduction for disasters caused by natural hazards, such as floods, cyclones, drought, or manmade disasters caused by industry. The design and development of training tools, such as an internet based training and knowledge management system and blended learning training methodology and the development of training materials are important activities under this project.

It gives me great pleasure to introduce these training modules to help trainees in developing knowledge and skill in specific thematic areas, to reduce the risk of disasters by use of databases and statistics.

I take this opportunity to express appreciation of the commitment of NIDM, Ministry of Home Affairs, Govt. of India, New Delhi, Dr. Klaus Röder and ifanos c&p Germany who extended their support and cooperation to this effort. I wish that such modules are used extensively by all stake holders across the country.

Dr. Satendra, IFS
Executive Director,
NIDM
New Delhi, September
2013

Foreword

Database on various aspects of hazards and vulnerability are important in planning and management of different stages of disaster management and emergency response. Data on hazards and disasters and Vulnerability/ development Indicators are being captured by various central, state government organizations, research and academic institutions as well as International organizations like UN agencies, Centre for Research on Epidemiology of Disasters (CRED) Munich–Re, Swiss–Re etc. However, there is less synergy in the data generated, with the needs of data users primarily due to gap in knowledge and skills of disaster database management at different levels. India has a robust statistical system to capture various statistics like environmental, health, housing, infrastructure, agriculture etc. however statistical system is yet to capture disaster damage and loss data systematically.

NIDM and GIZ Germany, under the aegis of Indo–German Environment Partnership Programme (IGEP) with Ministry of Environment and Forests, implemented a joint project entitled "Environmental Knowledge for Disaster Risk Management (ekDRM)" wherein development of case studies and training modules are the key activities. The modules on "Databases and Statistics for Disaster Risk Management" developed by Dr. Klaus Röder, Germany and the faculty members of NIDM is a praiseworthy effort.

The module covers extensively on various sources of hazard, disaster and resource database and how to access, collate and use the information for various phases of Disaster Risk Reduction. We hope the module will enhance the synergy between the data producers and the user community by developing the capacity of data producers to understand the requirement of the user community and also enhancing the knowledge of the disaster managers about the existing databases and their potential in managing disasters at various phases of DM Cycle. The training modules are tested at NIDM during the face to face training, blended learning and ToT programmes during 2010–13. I am sure, the module shall be of significant contribution in the capacity building and training within the country and outside. NIDM would welcome suggestions from the users of this module for improvements in future.

Contents

Module A: Disaster data/statistics – concepts and applications in disaster risk management

Introduction

Learning Unit A–1: Introduction to disaster management and vulnerability profile of India

Learning Unit A–2: Data for disaster management– broad categories

Learning Unit A–3: Data sources and accessibility

Module B: Introduction to statistical terminology

Learning Unit B–1: Introduction

Learning Unit B–2: The scenarios

Learning Unit B–3: Scenario of Bihar

Learning Unit B–4: Measures of central tendency and spread

Learning Unit B–5: More on measures of dispersion

Module C: Use Statistical Information on Recent and Past disasters

Learning Unit C–1: India and the international scenarios

Learning Unit C–2: Frequency and Impacts of Major disasters in India

Learning Unit C–3: Selected areas of disasters in India and their frequency and impacts

Module D: Understand ‘‘Cause and Effects’’ of disasters described by statistics

Learning Unit D–1: Introduction to the subject and test of hypothesis

Learning Unit D–2: Correlation – An introduction and relation to Disaster Risk Management

Glossary

List of acronyms

List of Figures

Figure A1.1:	Damaged school building in Sikkim	6
Figure A1.2:	Disaster management cycle	7
Figure A1.3:	Phases of Disaster Risk reduction cycle	8
Figure A1.4:	Earthquake hazard map showing faults and thrusts and earthquake with magnitude more than 5	14

Figure A1.5:	Flood hazard map showing area liable to floods.	15
Figure A1.6:	Wind and Cyclone Hazard map	16
Figure A1.7:	Landslide hazard zonation map	17
Figure A1.8:	Drought prone districts Identified by irrigation commission	18
Figure A1.9:	Thematic map showing concentration of MAH Units in Districts of India	20
Figure A3.1:	Hydrology data link at IMD website	29
Figure A3.2:	Climatology data at IMD website	29
Figure A3.3:	Preliminary Earthquake and Tsunami report in the IMD Website	30
Figure A3.4:	Home page of Central Water Commission	31
Figure A3.5:	Landslides home page in geological survey of India website	33
Figure A3.6:	Coverpage of Vulnerability Atlas of India	34
Figure A3.7:	Website of Decision Support Centre (DSC), NRSC	35
Figure A3.8:	Coverpage of Disastrous Weather Events Report, IMD	37
Figure A3.9:	Webpage of Drought Management Cell, Ministry of Agriculture	38
Figure A3.10:	Website of IDSP Programme, Ministry of Health	39
Figure A3.11:	Website of Chemical Accident Information and Monitory System, MoEF	40
Figure A3.12:	Homepage of DesInventar	42
Figure A3.13:	Homepage of EM–DAT, CRED	43
Figure A3.14:	Website of Census of India	48
Figure A3.15:	Website of DevInfo India	51
Figure A3.16:	Website of India Disaster Resource Network	53
Figure A3.17:	Website of State Disaster Resource Network	54
Figure B2.1:	People affected by disasters in India	67
Figure B3.1 :	Bihar administrative districts in 2011	72
Figure B3.2:	Bihar flood hazard map	73
Figure B3.3:	Website of the India Meteorological Department	74
Figure B3.4:	Bihar rainfall data from the India Meteorological Department website	75
Figure B4.1:	Composite of about 30 faces that were created to find out an average of good looks on the Internet.	80
Figure B4.2:	Skewedness	85
Figure B4.3:	Kurtosis	86
Figure B4.4:	Average monthly rainfall	87
Figure B4.5:	Screenshot 1 – IMD rainfall data	89
Figure B4.5:	Screenshot 2– IMD rainfall data	89
Figure B5.1:	Rainfall figures from the Bihar State (Box plot)	93

Figure B5.2: Screenshot of the StatistiXL menu on descriptive statistics 94
Figure B 5.3a, 3b and 3c – StatistiXL menu on descriptive statistics 95
Figure B5.4: Rainfall figures from the Muzzaffarpur District, Bihar State 96
Figure B5.5: Drought prone districts in Bihar 99
Figure C 1.1: Number of people killed and the costs of the (economic) damages due to major disasters in India between 1980 and 2004 116
Figure C1.2: Mortality Risk Index (MRI) for tropical cyclones 118
Figure C1.3: Vulnerability for tropical cyclones 120
Figure C1.4: Mortality Risk Index (MRI) for earthquakes 122
Figure C1.5: Vulnerability for earthquakes 122
Figure C1.6: Mortality Risk Index (MRI) for floods 124
Figure C1.7: Economic Risk Index (ERI) for floods 124
Figure C2.1: Time trend of reported natural disasters, 1975–2009 136
Figure C3.1: Cyclones in AP – damages and impacts 141
Figure C3.2: Gujarat earthquake hazard risk zonation:25 years return period estimated mean taluka peak ground acceleration (PGA) (in g). 146
Figure C3.3: Seismicity of India, 1990–2006 147
Figure C3.4: Standard operation procedure for managing flood hazards 153
Figure D1.1: Sample T–test 170
Figure D1.2: Sample T–test 170
Figure D2.1: Bihar: Rainfall and human casualties (1979–2008) 177

List of Tables

Table A1.1: Districts which are chronically prone to drought 19
Table A1.2: State–wise Number of MAH units and number of districts having MAH units 21
Table A3.1: Criteria for including an event in the database adopted in various international databases 41
Table B5–1: Open office syntax for statistical analysis 97
Table C2–1: Frequency and Impact of disasters 133
Table C2.2: Road traffic fatalities in India 134
Table C3.1: Ranking of disasters in different states 138
Table C3.2: Big earthquakes in India since 1990 148
Table C3.3: 10 year record of West Bengal floods 150

Module A

Disaster Data/Statistics Concepts and Applications in Disaster Risk Management

Contents

Module A

Introduction 1

1 Learning Unit A–1: Introduction to Disaster Risk Management, Hazard and Vulnerability Profile of India 5

1.1 Key terminologies 5

1.2 Typology of hazards and disasters 10

1.3 Hazards and vulnerability in India 12

2 Learning Unit A–2: Databases for Disaster Risk Management 23

2.1 About the Learning Unit 23

2.2 Potential uses of disaster databases 23

2.3 Databases for Disaster Risk Management – broad categories 24

2.3.1 Hazard data 24

2.3.2 Disaster data 25

2.3.3 Vulnerability indicators 25

2.3.4 Resource data 25

2.3.5 Miscellaneous data 26

3 Learning Unit A–3: Data Sources and Availability 27

3.1 About the learning unit 27

3.2 Hazard related statistics 28

3.2.1 India Meteorology Department (IMD) 28

3.2.2 Central Water Commission 31

3.2.3 Geological Survey of India 32

3.2.4 BMTPC – Vulnerability Atlas of India 34

3.2.5 National Remote Sensing Centre, Department of Space 35

3.3 Disaster database and statistics 37

3.3.1 Indian databases 37

3.3.2 International databases 41

3.3.3 CRED EM–DAT 42

3.3.4 Asian Disaster Reduction Centre (ADRC) 44

3.3.5 Australia: Emergency Management Australia Disasters Database 44

3.3.6 St. Lucia Disaster Matrix 45

3.3.7 Canada: Canadian Disaster Database 45

3.3.8 United States: SHELDUS 45

3.3.9 States: United States Storm and Hazard Database 46

3.3.10 United States: National Hazard Statistics 46

3.3.11 Philippines 47

3.3.12 Technological disasters: UNEP/APELL 47

3.3.13 Global Disaster Alert and Coordination System – GDAC 48

3.4 Data sets on development and vulnerability indicators 48

3.4.1 Census of India 48

3.4.2 Central Statistical Organization and National Sample Survey organization (NSSO) 49

3.4.3 Compendium of environmental statistics 50

3.4.4 Census Info/ DevInfo and emergency info 50

3.4.5 Planning Commission / State Planning Departments 51

3.4.6 National Informatics Centre 52

3.4.7 Environmental Information System (ENVIS) 52

3.4.8 Other sources 52

3.5 Resource inventory 53

3.5.1 India Disaster Resource Network 53

3.5.2 State Disaster Resource Network – Gujarat 54

3.5.3 Other Sources 54

3.6 Recent Initiatives 55

3.7 Gaps and challenges 56

4 Bibliography Module A 57

Introduction

Background

It is globally recognized that disaster losses are increasing exponentially. Natural and human induced disasters have been causing tremendous loss of life and property and damages to public and private infrastructure, eroding gains of development and also leading to mounting expenses on relief and rehabilitation. There are projections that disaster risks are likely to increase more under changing climatic conditions, increasing urbanisation and other factors like environmental degradation. Since later 1990s there is a paradigm shift in disaster management – from one of post disaster relief and reconstruction to a holistic management of disasters encompassing pre–disaster prevention, mitigation, preparedness, on–disaster response and post disaster relief, reconstruction and recovery. High Power Committee Report on Disaster Management, Chapter 7 gives special emphasis on knowledge networking, and developing of a resource database, although not discussed about disaster database/statistics (HPC Report, 2001). Globally this shift is reflected in the adoption of Yokohama Strategy of Safer World in 1994 and Hyogo Framework of Action in 2005. In India, the shift is articulated in the development of National Disaster Management Framework and reinforced through the enactment of Disaster Management Act 2005, National Disaster Management Policy, constitution of the National Disaster Management Authority and setting up of National Institute of Disaster Management. The National Policy on Disaster Management, Chapter 11 also envisages development of knowledge base for Disaster Risk Reduction (NDMA, 2009). Further the recent guidelines issued by National Disaster Management Authority also highlight the importance of carrying out detailed hazard, vulnerability and risk analysis while developing Disaster Management Plans at various levels (NDMA Guidelines).

Holistic management of disasters requires analytical data on hazards, risks and vulnerabilities of different natural and human induced disasters at all levels. Data on disaster occurrence, impact on people and its cost to countries are primary inputs to analyse the temporal and geographical trends in disaster impact. Disasters are real evidences to understand where the combinations of hazard occurrence, physical exposure and vulnerable conditions are accumulated in space and time. They also provide the basis for risk assessment processes to prioritize mitigation measures.

Disaster databases in our country are scattered and available with different department. At present data on different aspects of disasters are collected by various agencies, but mostly these remain confined to official files and sometimes periodicals and bulletins. There is no unified source of statistics on the hazard and disaster losses at the national/ sub–national levels in India.Similarly data on development indicators and vulnerability is also not available from any single source. Therefore a need has been felt for the development of a comprehensive database on disasters which would facilitate formulation of area specific disaster risk profile, assessment of long term impacts of disasters, development of policies, strategies and frameworks, preparation of proper plans for disaster preparedness and allocation of adequate funds for the prevention, mitigation and Disaster Risk Reduction (DRR).

India is country having robust statistical system with network of Central Statistical Organisation (CSO), and NSSO under Ministry of Statistics and Programme Implementation, State Level Directorate of Economics and Statistics with officers working up to district, sub–district levels and also with various departments at national and state levels. Indian Statistical System is capturing various hazard and development related information in the statistical handbooks already. States are already been publishing Environmental Statistics, Socio Economic Statistics etc. where the various indicators are captured. However the disaster databases are yet to find a place in these robust systems. A brainstorming workshop on the development of a National Disaster Statistical System was organized by the Central Statistical Organisation (CSO), Ministry of Statistics and Programme Implementation in collaboration with the National Institute of Disaster Management (NIDM), Ministry of Home Affairs on 27 April, 2007 at NASC, Pusa Complex, New Delhi. Various central government ministries and institutions like Indian Meteorological Department (IMD), Central Water Commission (CWC), Geological Survey of India (GSI), National Sample Survey Organization (NSSO), State Government Departments of Revenue and Disaster Managements and Directorates of Economics and Statistics (DES) attended the workshop. Two committees viz. Hazard Statistics and Disaster Statistics were constituted for facilitation of hazard and disaster statistics in July 2007. Directorate of Economics and Statistics at State level started including disaster related data also in the handbook, although the data is yet to be captured in the format finalised by the Central Statistical Organisation.

There is less synergy in the data generated, with the needs of data users, due to gap in knowledge and skills related to disaster database management, particularly the officials of statistics department. Similarly disaster management professionals are also not utilising the up to date data available in the statistical hand books and compendiums.

Aim of the Module

Aim of the module is to enhance the synergy between the data producers and the user community by developing the capacity of data producers to understand the requirement of the user community and also enhancing the knowledge of the disaster managers about the existing databases and their potential in managing disasters for Disaster Risk Reduction.

Target Group

The module is meant for the officials from State Department of Economics Statistics, State Disaster Management Authority, District Disaster Management Authority, other government and nongovernmental organisations and international agencies involved in Disaster Risk Management and also for researchers and training institutions like ATIs, SIRDs etc.

Structure of the Module A

The Module A has 3 learning units

Learning Unit A–1: Introduction to Disaster Risk Management, hazard and vulnerability profile of India

Learning Unit A–2: Data for disaster management– broad categories

Learning Unit A–3: Data sources and accessibility


For Databases
Statistics Management
and
for Managemtr
Disaster Dadabas

1

LEARNING UNIT A–1: INTRODUCTION TO DISASTER RISK MANAGEMENT, HAZARD AND VULNERABILITY PROFILE OF INDIA

Aim of the unit is to develop improved understanding of disaster related terminologies to officials from statistics department who are new to disaster management terminologies. This will enable the officials to collection and collation of statistics for DRM in systematic manner.

Key objectives of the learning unit are to:

- Develop understanding about the concepts are terminologies used in Disaster Risk Management;
- Generate awareness about the type of hazards and disasters;
- Introduce hazard and vulnerability profile of India.

1.1 Key terminologies

Hazard

Hazard is a potentially damaging physical event, phenomenon or human activity that may cause loss of life or injury, property damage, social and economic disruption or environmental degradation. Hazards are classified on the based on the origin i.e. Natural (geological, hydro–meteorological and biological), human induced process (environmental degradation and technological hazards). Hazards can be single, sequential or combined in their origin and effects. Each hazard is characterized by its location, intensity, frequency and probability.

Disaster

A disaster is a catastrophe, mishap, calamity in any area, arising from natural or manmade causes, which results in substantial loss of life or human suffering, damage destruction of, property, environment, and is of such a nature or magnitude as to be beyond the coping capacity of the community of the affected area. Impacts of disasters are

a) Loss of lives

b) Loss to property and infrastructure

c) Damage to livelihood

d) Economic losses

e) Environmental damage– flora & fauna

f) Sociological & psychological after effects

g) Civil strife

Figure A1.1: Damaged school building in Sikkim

Photo: SeedsIndia

Vulnerability

The conditions determined by physical, social, economic and environmental factors or processes, which increase the susceptibility of a community to the impact of hazards, are called vulnerability. Vulnerability can be due to the following factors

a) Geographic and environmental

b) Physical

c) Social

d) Economic

e) Systemic

Risk

Risk is the probability of harmful consequences or expected losses (deaths, injuries, property, livelihoods, economic activity disrupted or environment damaged) resulting from interaction between natural or human–induced hazards and vulnerable conditions. A simplified formula could read like this:

$$\text{Risk} \sim \frac{\text{Hazard x Vulnerability}}{\text{Capacity}}$$

Disaster Management

Disaster management is a continuous and integrated process of planning, organising coordinating, and implementing measures which are necessary for prevention of danger or threat of any disaster mitigation or reduction of risk, capacity building, preparedness, prompt response, severity or magnitude assessment, evacuation, rescue, relief, rehabilitation and reconstruction.

Disaster Risk Reduction

Disaster Risk Reduction is a systematic approach to identifying, assessing and reducing the risks of disaster. It aims to reduce vulnerabilities to disaster as well as dealing with the environmental and other hazards that trigger them.

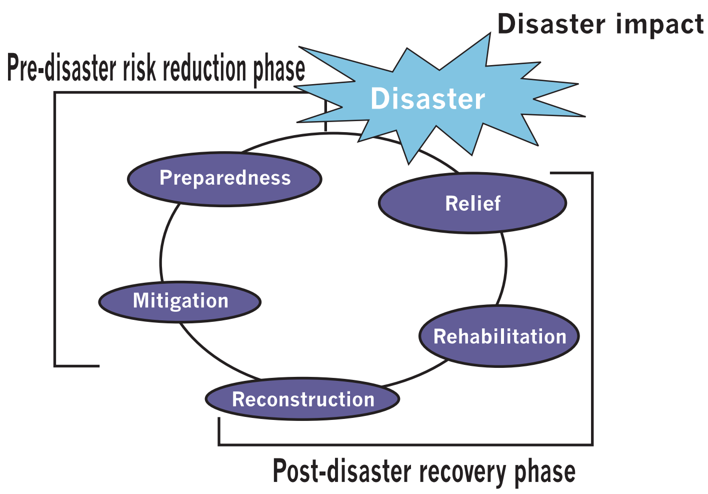


Figure A1.2: Disaster management cycle

Source: Own compilation

- "The conceptual framework of elements considered with the possibilities to minimize vulnerabilities and disaster risks throughout a society, to avoid (prevention) or to limit (mitigation and preparedness) the adverse impacts of hazards, within the broad context of sustainable development.

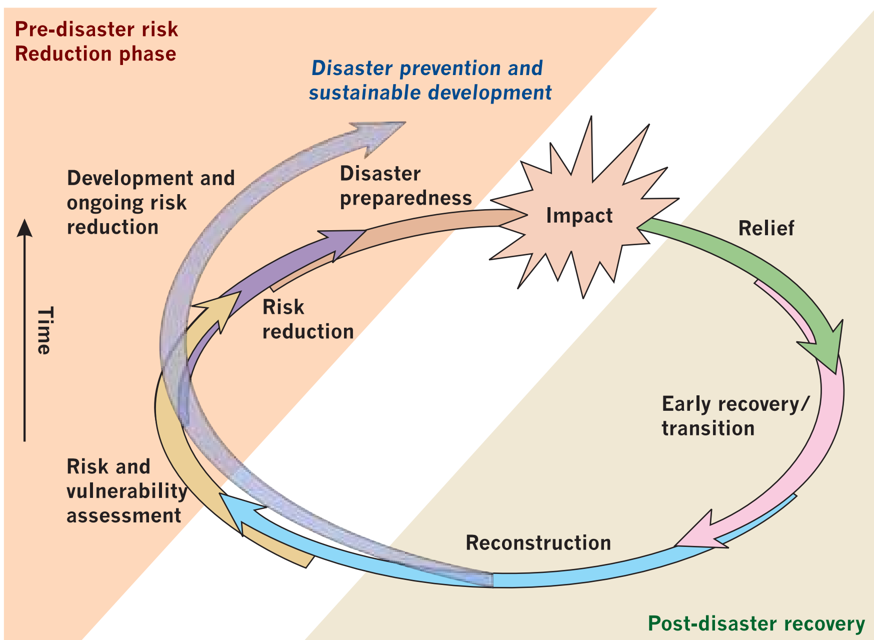


Figure A1.3: Phases of Disaster Risk reduction cycle

Source: Wisner B. et al. 2004

Prevention

Prevention means activities to avoid the adverse impact of hazards and means to check from turning into disasters. Examples: avoiding construction in seismically active areas, landslide prone areas and flood plains.

Mitigation

Mitigation means various structural and non–structural measures undertaken to limit the adverse impact of natural hazards, environmental degradation and technological hazards.

Preparedness

Preparedness includes measures taken in advance to ensure effective response to the impact of hazards, including the issuance of timely and effective early warnings and the temporary evacuation of people and property from threatened locations.

Response

Response includes specific actions taken immediately after a disaster to provide support to those affected. These activities are immediately initiated by the community itself and then by the district, state, national or up to international levels. These are actions and functions undertaken to face the disaster when it occurs. These include warning to vulnerable population, evacuation to avoid further damages, search and rescue, restoration of key infrastructure etc. A quick and effective response requires adequate planning and preparedness.

Relief

Relief are acts of helping or alleviating the conditions of persons who are suffering from the effects of disaster/calamity. A relief plan provides provisions of assistance or intervention during / immediately after a disaster to meet the basic needs of affected people.

Reconstruction

Reconstruction includes long term measures e.g. houses, livelihoods, infrastructures etc. It is capital intensive and henceneeds careful planning and community participation. The process also provides good opportunity to plan developmental activities which are more robustand disaster resilient.

1.2 Typology of hazards and disasters

High Powered Committee on Disaster Management identified 33 different types of disasters affecting various parts of India. They are classified under five groups based on the origin or the cause.

I. Water and climate related disasters

1. Floods
2. Cyclones
3. Tornadoes
4. Hailstorm
5. Cloud burst
6. Heat wave and cold wave
7. Snow avalanches
8. Droughts
9. Sea erosion
10. Thunder and lightning
11. Tsunami (not in the HPC list)

II. Geologically related disasters

1. Landslides and mudflows
2. Earthquakes
3. Dam failures/ dam bursts
4. Mine fires

III. Chemical, industrial and nuclear

1. Chemical and industrial disasters
2. Nuclear disasters

IV. Accident related disasters

1. Forest fires
2. Urban fires
3. Mine flooding
4. Oil spill
5. Major building collapse
6. Serial bomb blasts
7. Festival related disasters
8. Electrical Disasters and Fires
9. Air, Road and Rail Accidents
10. Boat Capsizing
11. Village Fire

V. Biologically related disasters

1. Biological disasters and epidemics
2. Pest attacks
3. Cattle epidemics
4. Food poisoning

Disasters have been classified into three groups based on their intensity or magnitude (Level 1, Level 2 and Level 3). The classification has been made with the philosophy that for Level 1 emergency, the District Emergency Response Group would be able to take control of the situation. For a Level 2 scenario the State Emergency Response Group would be activated and for a Level 3 disaster, the National Emergency Response Group comes into the picture (adapted from HPC Report, 2000).

1.3 Hazards and vulnerability in India

A disaster occurs when hazard interacts with vulnerable conditions. For example, if an earthquake (hazard) occurs, a structurally safe building will withstand the shock (resistant), but a house/ building (vulnerable) may collapse; creating a disaster for the dwellers. Vulnerability could be due to the human related factors or natural features. The human related factors that increase vulnerability of India could be intended or unintended, and include poverty, corruption, illiteracy, land use pattern, technological misuse, and terrorism. Poor land use planning and inconsistent emergency management systems lead to vulnerability to floods, drought, cyclones, earthquake, heat and cold waves, and landslides.

As mentioned, India has a highly diversified range of natural features. Its unique geo–climatic conditions make the country among the most vulnerable to natural disasters in the world. Disasters occur with very high frequency in India and while the society at large has adapted itself to these regular occurrences, the economic and social costs continue to mount year after year. It is highly vulnerable to floods, drought, cyclones, earthquakes, landslides etc. Almost all parts of India experience one or more of these events.

Many regions in India are highly vulnerable to natural disasters on account of geological conditions. About 59 % of the total area of the country is vulnerable to seismic damage of buildings in varying degrees (BMTPC, 1997 &2006). The most vulnerable areas, according to the present seismic zone map of India, are located in the Himalayan and sub–Himalayan regions, Kutch and the Andaman and Nicobar Islands, which are particularly earthquake hazard prone (Figure A1.4). Over 8% Indian area i.e. 40 million hectares is prone to floods (Figure A1.5), and the average area affected by floods annually is about 8 million hectares. Of the nearly 7,500 kilometres long coastline, approximately 5,700 kilometres is prone to cyclones (Figure A1.6), and 68% area is susceptible to drought. Irrigation commission report indicates that there are 102 districts spread across 15 states of India which are chronically prone to drought (see Table A1.1 and Figure A1.8). 15% of the Geographical area covering Himalayan range, Western and Eastern Ghats are prone to Landslides (Figure A1.7)

Besides the major disasters, the country is prone to several other disasters chronically i.e. heat wave, cold wave, hail storms, thunderstorms and lightening, coastal and sea erosion, biological hazards like epidemics and pest attacks and so on. The northern parts of India specially the hilly regions and the adjoining plains experience cold wave conditions. The cold waves mainly affect the areas to the north of 20 degree N but in association with large amplitude troughs, cold wave conditions are sometimes reported from States like Maharashtra, Andhra Pradesh and Karnataka as well. The maximum numbers of cold waves occurred in Jammu & Kashmir (211 times) followed by Rajasthan (195) and

Uttar Pradesh (127) during 1900–1999 (Dey et al., 2005). Maximum number of heat waves occur over East Uttar Pradesh (134) followed by Punjab, east Madhya Pradesh and Saurashtra & Kutch in Gujarat. Month wise, maximum number of heat waves occurred during the month of June.

Climate change is likely to pose new threats in the coastal region. Coastal region is vulnerable to inundation from accelerated sea level rise like the Lakshadweep archipelago, the east coast of India, with its lower coastal slopes and higher cyclone frequency, will suffer from increased storm surge damage and the belt between approximately12°N and 18°N on the west coast appears to be less vulnerable however the region to the south of this belt is likely to experience increased coastal erosion (Shetye et al., 1990)

Disasters are no longer limited to natural catastrophes. There are 1861 Major Accident Hazard Units in India. There are 298 districts in India having major accident hazard units which are spread across 28 states and UTs (Table A1.2 and FigureA1.9).

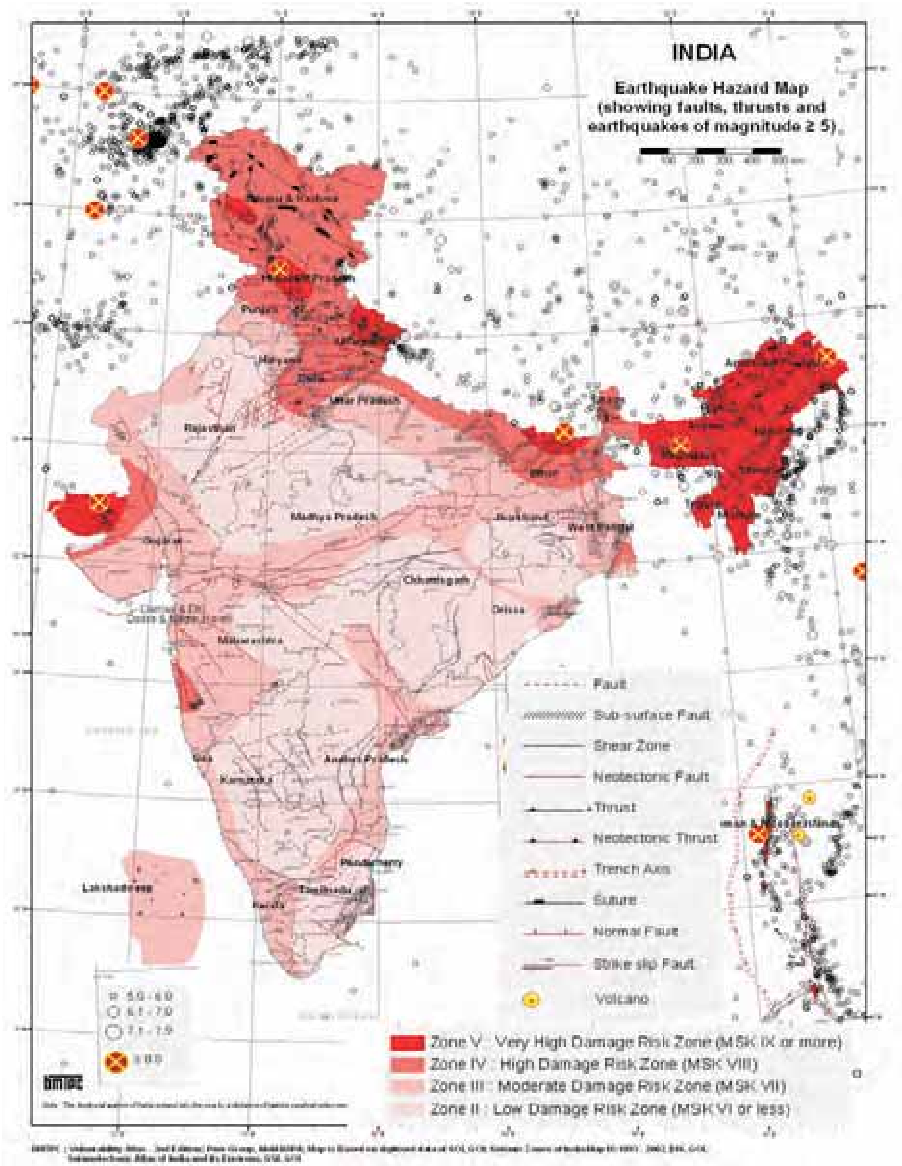


Figure A1.4: Earthquake hazard map showing faults and thrusts and earthquake with magnitude more than 5.

Figure A1.5: Flood hazard map showing area liable to floods.

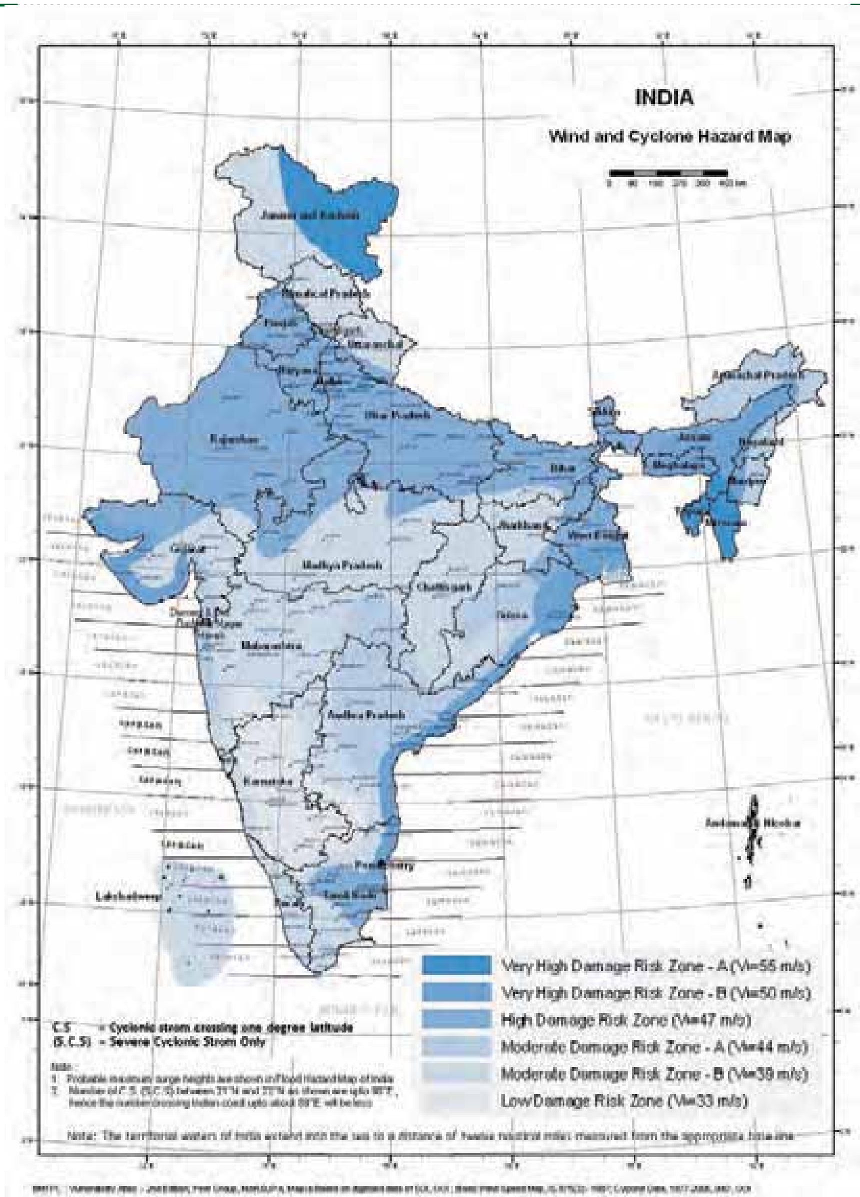


Figure A1.6: Wind and cyclone hazard map of India

Man–made emergencies also cause disasters in terms of fatalities and economic losses.

The Vulnerability Atlas of India 1997 shows that there are 169 districts prone to multi hazards considering only floods, earthquake and cyclone. As per the revised atlas 2006, there are 241 district covering 20 states which are prone to multiple hazards.

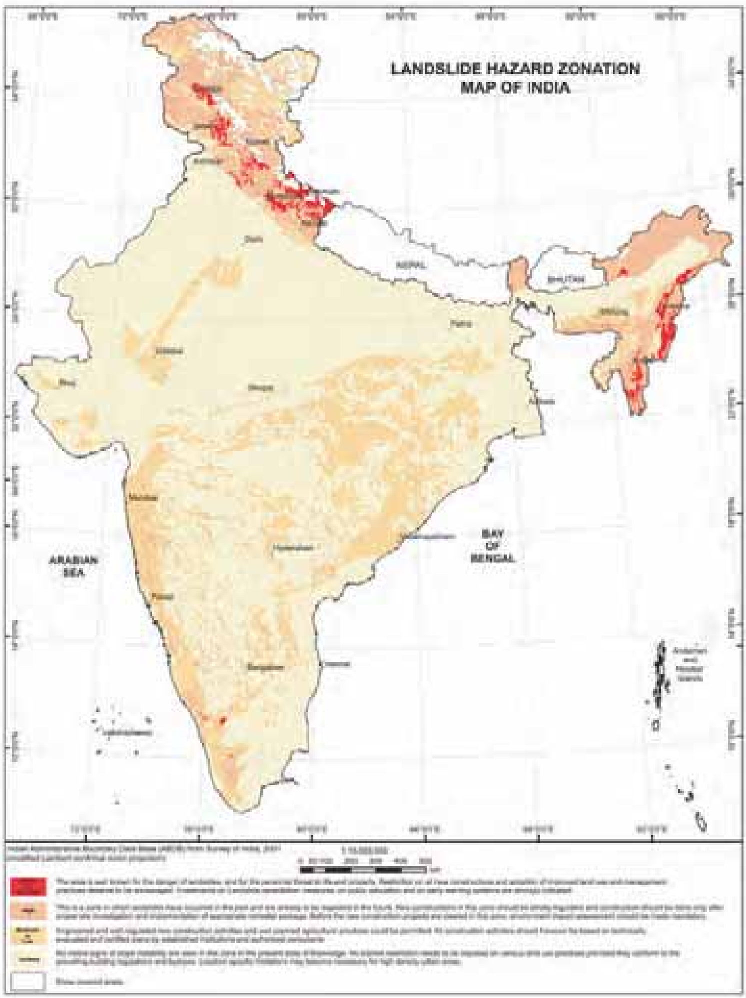


Figure A1.7: Landslide hazard zonation map

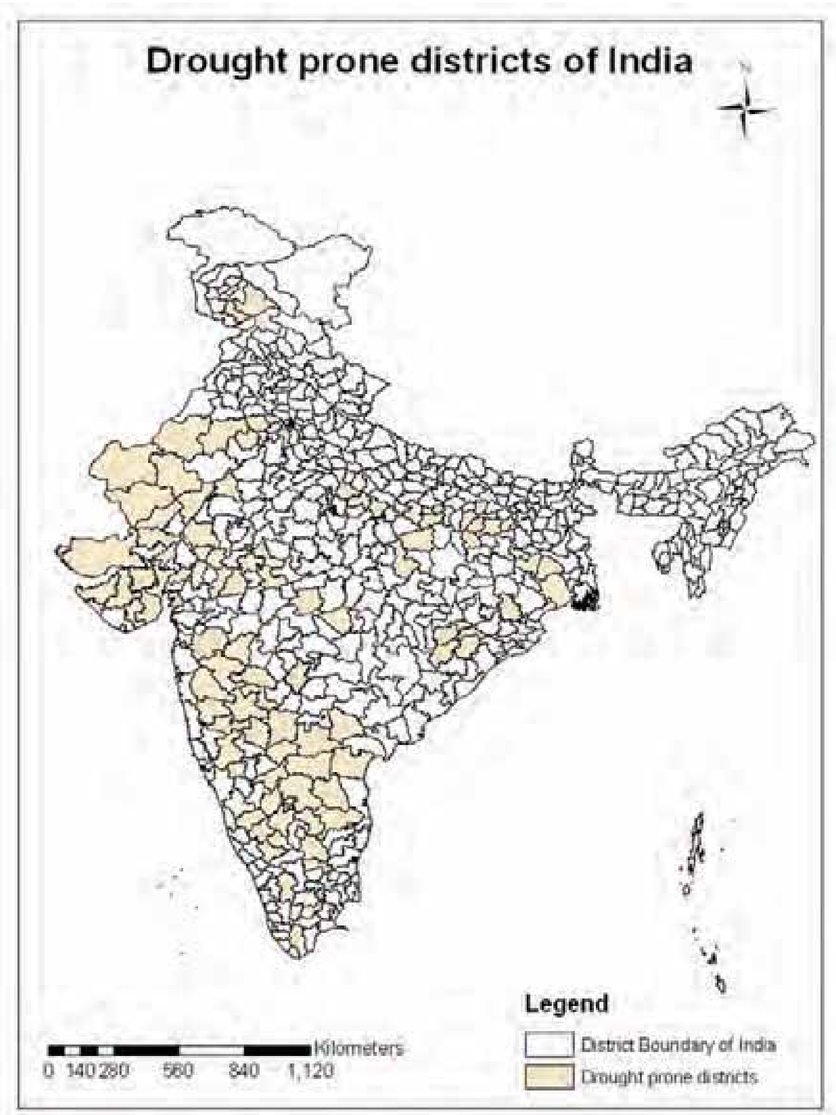


Figure A1.8: Drought prone districts Identified by irrigation commission

Source: National Drought Manual, 2009

Table A1.1: Districts which are chronically prone to drought

State	Districts
Andhra Pradesh	Anantpur, Chittoor, Cuddapah, Hyderabad, Kurnool, Mehboobnagar, Nalgonda, Prakasam
Bihar	Munger, Nawadah, Rohtas, Bhojpur, Aurangabad, Gaya
Gujarat	Ahmedabad, Amreli, Banaskantha, Bhavnagar, Bharuch, Jamnagar, Kheda, Kutch, Mehsana, Panchmahal, Rajkot, Surendranagar
Jammu and Kashmir	Doda, Udhampur
Karnataka	Bangalore, Belgaum, Bellary, Bijapur, Chitradurga, Chickmagalur, Dharwad, Gulbarga, Hassan, Kolar, Mandya, Mysore, Raichur, Tumkur
Madhya Pradesh	Betul, Datia, Dewas, Dhar, Jhabua, Khandak, Shahdol, Shahjapur, Sidhi, Ujjain
Maharashtra	Ahmednagar, Aurangabad, Beed, Nanded, Nashik, Osmanabad, Pune, Parbhani, Sangli, Satara, Solapur
Orissa	Phulbani, Kalahandi, Bolangir, Kendrapada
Rajasthan	Ajmer, Banswada, Barmer, Churu, Dungarpur, Jaisalmer, Jalore, Jhunjunu, Jodhpur, Nagaur, Pali, Udaipur
Tamil Nadu	Coimbatore, Dharmapuri, Madurai, Ramanathapuram, Salem, Tiruchirapali, Tirunelveli, Kanyakumari
Uttar Pradesh	Allahabad, Banda, Hamirpur, Jalana, Mirzapur, Varanasi
West Bengal	Bankura, Midnapore, Purulia
Jharkhand	Palamau
Chhattisgarh	Khargaon

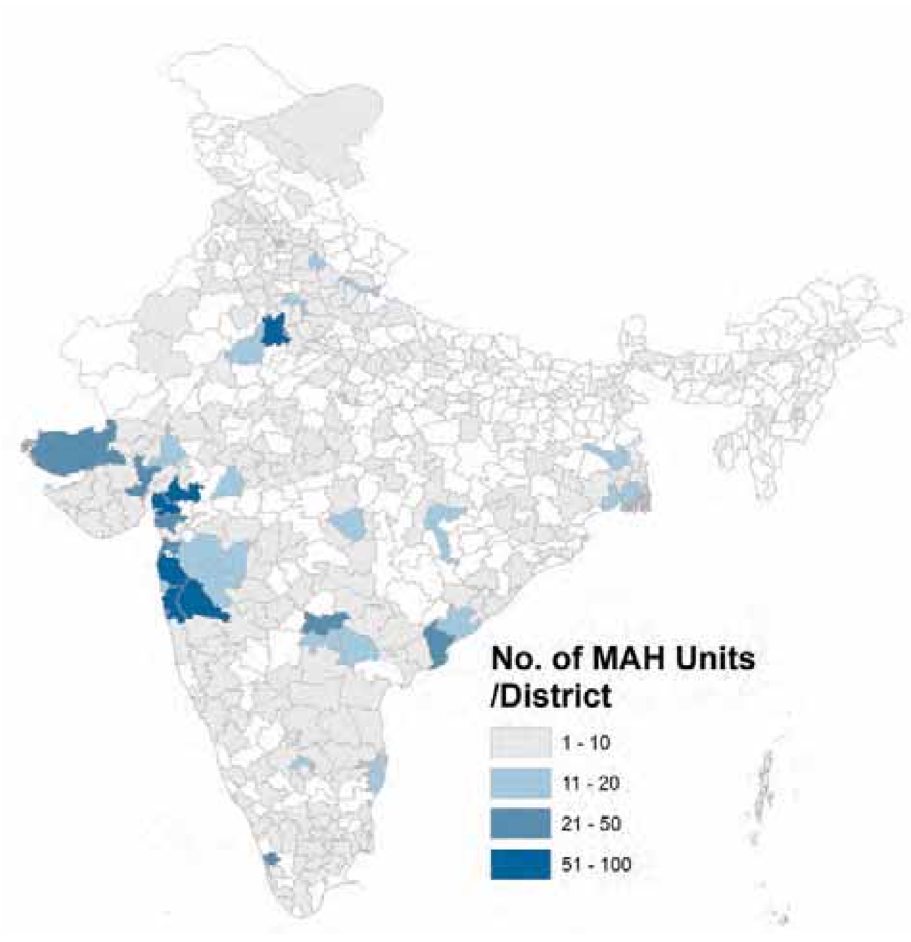


Figure A1.9: Thematic map showing concentration of MAH Units in Districts of India

Source: This map is based on the list if MAH Units received from HSMD in Dec 2012

Table A1.2: State–wise Number of MAH units and number of districts having MAH units

State	Nos. of MAH Units	Dist. with MAH units
A&N (2010)	3	1
A.P. (2012)	158	21
Arunachal Pradesh (2007)	No MAH Unit	
Assam (2012)	26	10
Bihar (2012)	11	4
Chandigarh (2010)	No MAH Units	
Chhattisgarh (2012)	19	4
Daman & Diu (2010)	No MAH Units	
Dadra & N.H. (2010)	No MAH Units	
Delhi (2011)	18	7
Goa (2011)	16	2
Gujarat (2011)*	390	23
Haryana (2010)	58	13
H.P. (2012)	8	4
J&K (2010)	13	6
Jharkhand (2011)	13	5
Karnataka (2012)	73	16
Kerala (2011)	46	9
Lakshadweep (2009)	No MAH Units	
M.P. (2012)	72	21
Maharashtra (2011)*	369	27
Manipur (2012)	2	1
Meghalaya (2010)	2	1
Mizoram (2010)	No MAH Units	
Nagaland (2012)	2	1
Orissa (2012)	23	13
Pondicherry (2012)	3	2
Punjab (2012)	64	14
Rajasthan (2011)	109	14
Sikkim (2009)	No MAH Units	
Tamil Nadu (2009)	118	24
Tripura (2012)	2	1
Uttar Pradesh (2010)	120	38
Uttarakhand (2011)	38	4
West Bengal (2012)	85	12
	1861	298

2

LEARNING UNIT A–2: DATABASES FOR DISASTER RISK MANAGEMENT

2.1 About the Learning Unit

In the previous unit we have learned about the basic concepts and terminologies used in disaster management, phases of disaster management cycle and vulnerability profile of India. Now let us group the data/ statistics which can be used for disaster management in various phases of disaster management cycle.

Aim of this unit is to introduce various types of data required for Disaster Risk Management and their uses during different phases of Disaster Management.

Objectives of the Learning Unit

- Generate awareness about the potential uses and users of databases
- Identify and categorise different databases and statistics required for DRM

2.2 Potential uses of disaster databases

Disasters are the instruments for measuring the existing hazards, vulnerability and risk in space and time.

Some of the potential uses of disaster databases are as follows:

Pre–disaster phase

- To carry out hazard, vulnerability, capacity and risk analysis. Developing disaster risk indexing system that tracks the patterns of disaster risk – spatially and temporally.

- Validation of models developed based on Indicators.
- Develop a policy advocacy tool for drawing attention to disaster issues for prioritizing mitigation measures
- Analyse how development policies and practices have enhanced or reduced disaster risk of area or a community. (Monitoring and Evaluation).

During disaster and post– disaster phase

- Response planning: access to reliable information/ vulnerability analysis can improve the early warning and effective dissemination and also help in better co–ordination between different agencies providing relief and recovery assistance.
- Situation analysis and reporting: geographical scale of impact, losses and population affected for preliminary analysis.
- Damage assessment: Statistical database on the sectoral damages and geographic distribution (segregated and aggregated)
- Recovery framework: can propose appropriate mitigation measures through reconstruction and rehabilitation strategies based on vulnerability data.

2.3 Databases for Disaster Risk Management – broad categories

2.3.1 Hazard data

Hazard data include various hydrological, geological, meteorological and manmade threats. Data base of heavy rainfall, extreme temperature, seismic activity, depressions, location of hazardous industries etc. are examples of hazards. Every hazard is characterized by a frequency, intensity and duration. Data on hazards are used for mapping and identifying the areas prone to heavy rain fall, seismicity, extreme weather events etc. This data is a key input for development planning. Hazard data is used for developing early warning systems, risk assessment and for disaster management planning.

2.3.2 Disaster data

Data of all the events happened deals with the death, injuries, damages and losses. Refer to definition of Disaster in Unit 1.

Some of the potential uses of disaster databases are as follows:

Disaster data can be used for developing disaster risk indexing system that tracks the patterns of disaster risk – spatially and temporally, developing policy advocacy tool for drawing attention to disaster issues for prioritizing mitigation measures and also for analysing how development policies and practices have enhanced or reduced disaster risk of an area or a community. During or in the post– disaster phase situation analysis and reporting i.e. geographical scale of impact, losses and population affected for preliminary analysis, response planning; access to reliable information can improve co–ordination between different agencies providing relief and recovery assistance, damage assessment; statistical database on the sectoral damages and geographic distribution (segregated and aggregated)and recovery framework; while developing recovery framework appropriate mitigation measures can be proposed through reconstruction and rehabilitation strategies based on data.

2.3.3 Vulnerability indicators

Vulnerability indicators include various Physical and Socio economic factors which convert hazard into a disaster: Unsafe buildings, poor infrastructure, high population density, poor planning and enforcement mechanisms etc. Without understanding the vulnerability of the affected population or the population likely to be affected by a disaster appropriate measures cannot be adopted for reducing disaster risk. Factors leading to increasing susceptibility may be identified using various demographic, social, economic, physical infrastructure, Land use and land cover and exiting policy initiatives. This is complex methodology and requires wide range of datasets

2.3.4 Resource data

These data include material resources and human resources and skills. These data also can be classified based on the utility, availability (place and time), channels for delivering cost and time requirement. Such resource data is extremely useful in responding to emergencies. The data can be used for disaster management planning, preparedness planning etc.

2.3.5 Miscellaneous data

These data include information not directly related to disasters for e.g. infrastructure data or agricultural data to name just a few. This information is the most frequent and abundant and often readily available. The task with this information is to select the information important for Disaster Risk Management and how to use and analyse data for the purpose of disaster management. However a database with diligent presentation of relation between causes and disaster effects or the comparison of successful and less effective relief and rescue campaigns based on miscellaneous data will be helpful for improving the Disaster Management Systems.

3

LEARNING UNIT A–3: DATA SOURCES AND AVAILABILITY

3.1 About the learning unit

Previous unit we have learned the broad categories of datasets which are used for Disaster Risk Management and their applications. Information and Data on hazards and disasters and vulnerability/ development Indicators are being captured by various central, state government organizations, research and academic institutions as well as International organizations like UN agencies, CRED etc. Major sources of hazard, disaster and vulnerability indicators and sources of information and organizations / institutions generating those datasets in India are given below. Now let us explore the sources of the existing data sets at state, national, regional and international levels. At the end of this unit participants shall identify the sources of data, availability and spatial and temporal scales for the various hazards and disasters at the state level.

Aim of the unit is to generate awareness about the sources of data and how to access the data.

Objectives

- Identify the sources of data on various hazards.
- Identify the sources of data vulnerability and development indicators
- Identify the existing databases on disasters and analyse strengths and gaps
- Introduce the sources of disaster damage data in India and mechanisms of capturing
- Introduce to resource databases for DRM in India

3.2 Hazard related statistics

3.2.1 India Meteorology Department (IMD)

www.imd.gov.in

IMD brings out hazard and disaster data products and publications regularly for the benefit of user agencies since its establishment in 1876. A variety of climatological data are regularly supplied by the National Data Centre, Pune and other departmental offices to Central and State Governments, Universities, Research Institutes, Public undertakings and Private sectors. The information supplied is used for town planning, setting up of new industry, port installations, installation of high towers, bridges and other structures, operation of multipurpose hydel projects, proper water and power management, defence operations in inaccessible regions, calibration of defence equipment, environmental studies, non–conventional energy source, lay–out of run–ways and disaster management planning. IMD has also undertaken studies to brief Central and State Government authorities regarding prospects of monsoon, utilisation of wind–power and off–shore drilling operations etc. The department has published the first **Climatological Atlas of India** in 1906, Climatological Atlas of India (Abridged) in 1971, Rainfall Atlas of India in 1971 and reprinted in 1972, Climatological Atlas of India Part–A (Rainfall) in 1981 and Marine Climatological Atlas (1961–1990), Climatological features of drought incidences in India (2005) and Monthly and annual normals of rainfall and of raingauges for IMD observatories (1901–1950) and High Resolution Grided Rainfall Data (1951–2003) for the Indian region (2006). Apart from this IMD has come up with reports and atlas for Climate of Maharashtra (1971 and revised in 2005), Andhra Pradesh (1973), Climate of Kerala (1986), Karnataka (1984), Rajasthan (1988), Climate of Uttar Pradesh (1989), Haryana and Union territories of Delhi and Chandigarh (1991), Gujarat and Diu, Daman, Dadra & Nagar Haveli (1995), Punjab (1996), Tamil Nadu (including Pondicherry) and Karaikal (2000), Orissa (2002).

IMD is also preparing detailed maps of rainfall, temperature (maximum and minimum), cloud cover, relative humidity and analyse this information with prevailing crop conditions and an Agromet Advisory Bulletin is prepared and disseminated to users. Based on rainfall, temperature, soil moisture and evaporation, various indicators of meteorological drought indicators have been developed by researchers.

Some of the products by IMD:

1. All India Weather Report with average temperature (max and min) and rainfall for all the cities of India during the period 1901–2000. http://www.imd.gov.in/doc/climateimp.pdf

2. Hydrology data

Rainfall data, 100 Years (1901–2000) Monthly Rainfall Data Series for Districts, States, Met– Subdivisions and all India is available with Hydrometeorology Division, Office of the Additional Director General of Meteorology (Research), and Pune.

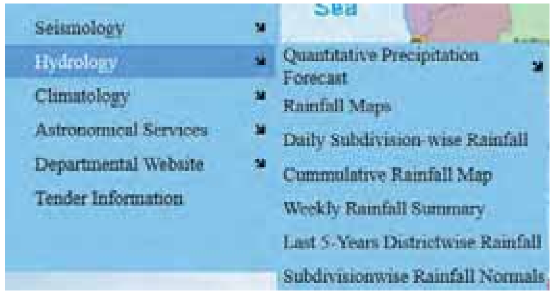


Figure A3.1: Hydrology data link at IMD website

http://www.imd.gov.in/section/nhac/dynamic/–advertisement_district_rf_series.pdf

Quantitative precipitation forecasts for cities and towns, rainfall maps, daily subdivision wise rainfall, cumulative rainfall maps, Weekly rainfall summary, last 5 years district wise rainfall, sub division wise rainfall normal etc. are the other data generated and published by IMD. Most of these data are available online in the public domain and can be accessed.

3. Climate data

IMD produces wide range of climate related data which includes, climate profile, climatologically summaries of daily temperature, climate normals, monthly mean maximum & minimum temperature and monthly total rainfall of important stations for the period 1901–2000 and so on. These data sets were accessible online at http://www.imd.gov.in/.

Figure A3.2: Climatology data at IMD website

4. Seismology data

IMD also generates preliminary reports of earthquakes and tsunami.

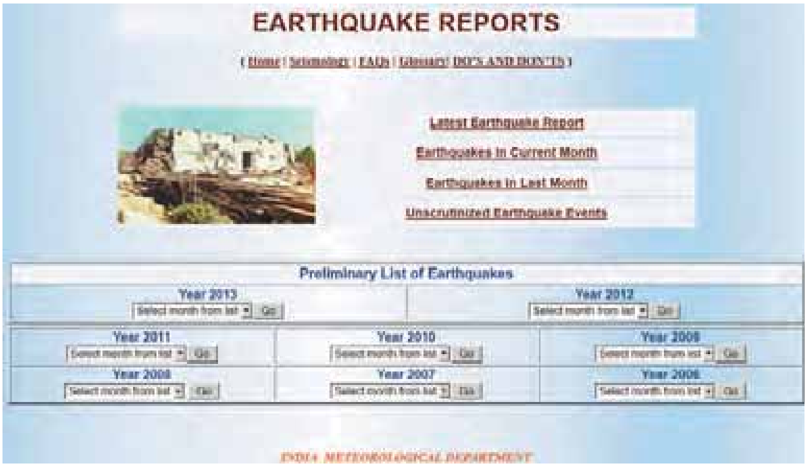


Figure A3.3: Preliminary Earthquake and Tsunami report in the IMD Website

India Meteorological Department is maintaining a country wide National Seismological Network (NSN), consisting of a total of 82 seismological stations, spread over the entire length and breadth of the country. This includes: a) 16–station V–SAT based digital seismic telemetry system around National Capital Territory (NCT) of Delhi, b) 20– station VSAT based real time seismic monitoring network in North East Region of the country and (c) 17–station Real Time Seismic Monitoring Network (RTSMN) to monitor and report large magnitude under–sea earthquakes capable of generating tsunamis on the Indian coastal regions. The remaining stations are of standalone/ analogue type. IMD is having state–of–the art facilities for data collection, processing and dissemination of information to the concerned user agencies. Seismology division supplies earthquake data and seismicity reports of specific regions to various user agencies such as, insurance companies, industrial units, power houses, river valley projects etc. on payment basis. Seismological data and earthquake related information is also supplied to agencies dealing with relief and rehabilitation measures, earthquake disaster mitigation and management related matters, seismic zoning etc. Earthquake data is shared with various scientific, academic and R&D institutions for research purposes. Towards early warning of tsunamis, real–time continuous seismic waveform data of three IMD stations, viz., Portblair, Minicoy and Shillong, is shared with global community, through IRIS (Incorporated Research Institutions of Seismology), WashingtonD.C., USA. Preliminary report gives information on earthquake epicentre, magnitude, date and time and depth and region. Recent earthquake information from the year 2006 to 2013 is available online and last 100 years data available with IMD Seismology Division, New Delhi.

http://www.imd.gov.in/section/seismo/dynamic/welcome.htm.

3.2.2 Central Water Commission (www.cwc.nic.in)

Central Water Commission at present operates nationwide network of 945 hydrological observation stations. Out of these 945 stations, 246 are gauge sites, 282 are gauge and discharge sites, 115 are gauge discharge and water quality sites, 41 are gauge, discharge and silt sites, while the remaining 261 are gauge, discharge, silt and water quality sites. The river management wing of CWC is responsible for, collection, compilation, storage and retrieval of hydrological and hydro–meteorological data including water quality monitoring, formulation and issue of flood forecast on all

Figure A3.4: Home page of Central Water Commission major flood prone rivers and inflow forecasts for selected important reservoirs, providing guidance to states in technical matters on different aspects of river and flood management in the country and regulation of multipurpose reservoirs, river morphology studies, techno–economic appraisal of various flood management schemes received from the State Governments, providing advice to coastal states on issues related to coastal erosion problems including preparation of National Coastal Protection Project (NCPP) for coastal protection works, survey and investigation of water resources development projects in India and neighbouring countries and international co–operation with neighbouring countries in the field of flood forecasting.Flood forecasting service of CWC cover almost all major flood prone inter–state river basins of India. At present there are 145 level forecasting stations on major rivers and 27 inflow forecasting stations on major dams/barrages. It covers 9 major river systems in the country, including 65 river sub–basins pertaining to 15 states viz. Andhra Pradesh, Assam, Bihar, Chhattisgarh, Gujarat, Haryana, Jharkhand, Karnataka, Madhya Pradesh, Maharashtra, Orissa, Tripura, Uttaranchal, Uttar Pradesh & West Bengal and one union territory Dadra & Nagar Haveli and the National Capital Territory of Delhi. Normally forecasts are issued 12 to 48 hours in advance, depending upon the river terrain, the location of the flood forecasting sites and base stations.

CWC has been issuing Daily Flood Bulletins and Special Flood Bulletins during the flood season every year based on the information collected from affected State Governments and its own field formations. The organization is providing information support by collecting, storing and disseminating statistical data, creation and maintenance of computerized data bank on water & related data, documentation of water resources and related data including data on financial aspects of irrigation projects and hydrological data for non–classified basins in the form of publications and undertaking analysis of data for providing decision support. Key publications on water related hazards are National Register of Large Dams (2009), Integrated Hydrological Data Book, Water & Related Statistics and **Preliminary Consolidated Report on Effect of Climate Change (2008)**. Water and related statistics gives information on seasonal average runoff at CWC sites for all the river basins, seasonal average observed suspended load, flood damages and heavy rains in India 1953–2002 and flood forecasting performance which are relevant for flood management. Details of CWC data are available at www.cwc.nic.in

3.2.3 Geological Survey of India

GSI has developed and published (Special Publication No. 71) **An Inventory of Landslide Incidences** from NW Himalayas, Eastern Himalayas & North–eastern states in the year 2005.Glaciology Division of GSI had initiated modern methods in the glaciological studies which included study of glacier regimen, mass balance, hydrometry etc. glacial and periglacial geomorphology, snow cover studies

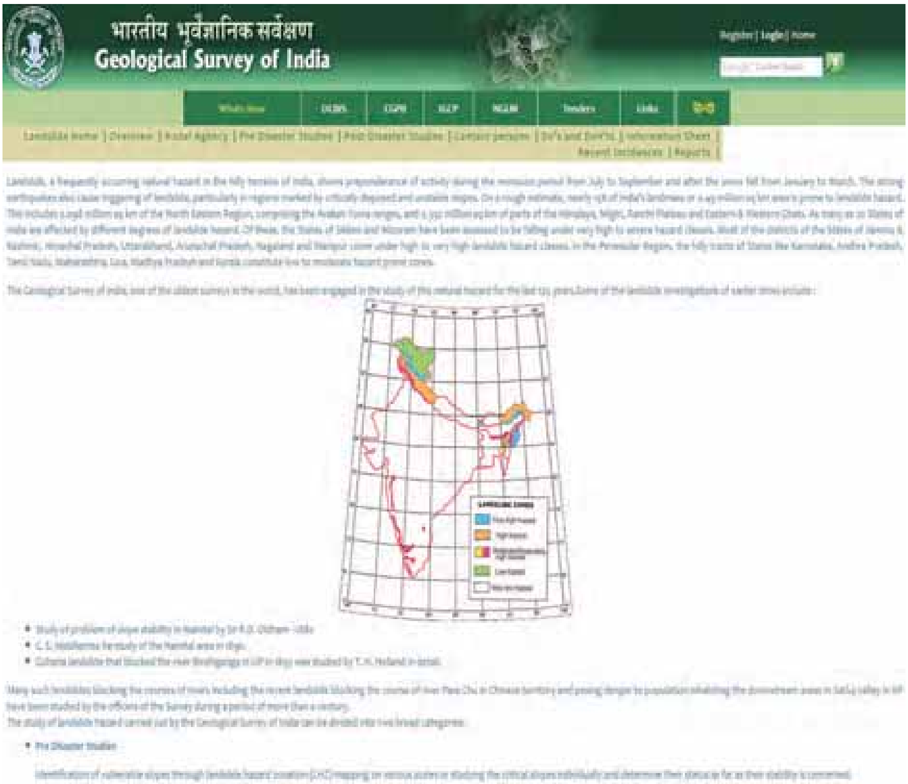


Figure A3.5: Landslides home page in geological survey of India website

Source: http://www.portal.gsi.gov.in/portal/page?_pageid=127,671641&_dad=portal&_schema=PORTAL (last access 17.9.2013)

and inventory of glaciers in the individual basins. Detailed glaciological studies were initiated at Gara glacier (Tirung Khad basin), Kinnaur district in H.P. in the year 1974 and subsequently Neh Nar, Rulung, J&K, GorGarang, H.P. Tipra Bank, Dunagiri, ShauneGarang, H.P., Harmukh, J&K., Triloknath, H.P. were studied. Presently glaciological studies at Hamtah glacier, H.P. are in progress. A preliminary estimate of the Indian Himalayan glaciers was published in World Glacier Inventory Status 1988 whereas the detailed catalogue of the glaciers was recorded in Inventory of the Himalaya Glacier as GSI Spl. Pub. No.34 (1999). Secular movement of the glaciers gives an insight into the past climate and the prevailing climate of a region. The hazards of higher reaches in the Himalaya in from of avalanches were assessed in Sind valley, J&K, Kedarnath area and Dhauliganga valley in U.P. The hazardous area in the lower reaches of the Himalaya delineated while mapping the glacio–geomorphic features. Possible risk to engineering structures in Khardung La, Hung dam and in Batkot area J&K, Kedarnathtownship, Uttarakhand and power project in DhauligangaValley, Uttarkhand were also assessed by the division.

3.2.4 BMTPC – Vulnerability Atlas of India

The first Vulnerability Atlas has been published by the Building Material Technology Promotion Council (BMTPC), Ministry of Urban Development and Poverty Alleviation in the year 1997 and the atlas was revised in 2006. The Vulnerability Atlas gives state–wise hazard maps and district–wise damage risk tables for all States and U.Ts. This gives information on all vulnerable areas that have been mapped indicating the expected intensities of natural events earthquakes, cyclones and floods and landslides (included in 2nd Edition released in 2006) that can occur in different areas. It also gives the expected risk levels of different types of housing stock existing in each district. The information and data given in the Atlas can be utilised for preparing damage scenarios for individual districts of states for a possible occurrence of a hazard, and to formulate suitable strategy to prepare at macro–level. The information given in the Atlas will help the planning agencies, the state and district administrations and the community at Panchayat levels in raising the level of awareness about the possible risk in the disaster prone areas and the measures required for preparedness and mitigation on scientific and realistic basis. Details of the publications are available at www.bmtpc.org.

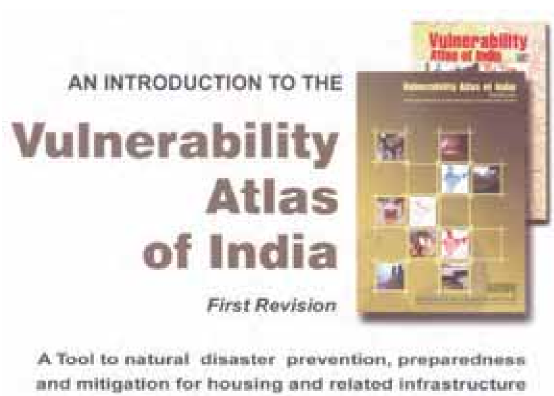


Figure A3.6: Coverpage of Vulnerability Atlas of India

3.2.5 National Remote Sensing Centre, Department of Space

ISRO/DOS is playing a vital role in supporting the disaster management activities, by providing space as well as aerial remote sensing based services and products. The Decision Support Centre (DSC) is established at National Remote Sensing Centre (NRSC) as part of Disaster Management Support Programme of Department of Space (DOS), for working towards effective management of disasters in India. Using satellite data from Indian Remote Sensing Satellite (IRS) System and from foreign satellites, the impact of floods in the country is being assessed for the past one decade operationally.

Figure A3.7: Website of Decision Support Centre (DSC), NRSC

Source: http://dsc.nrsc.gov.in/DSC/index.jsp

Services include near real time flood mapping and monitoring, flood damage assessment, flood hazard zone mapping, river bank erosion mapping, mapping changes in the river course. **National Agricultural Drought Assessment and Monitoring System (NADAMS)** is being developed by the Department of Space for the Department of Agriculture and Co–operation in 1989. **NADAMS** provides near real–time information on prevalence, severity level and persistence of agricultural drought at state/district/sub–district level. Currently, the project covers 13 states of India, which are predominantly agriculture based and prone to drought situation. Agricultural conditions are monitored at state/district level using daily observed coarse resolution (1.1 km) NOAA AVHRR data for 11 states. Moderate resolution data from Advanced Wide Field Sensor (AWiFS) sensor of resources at 1 (IRS P6) of 56 m and Wide Field Sensor (WiFS) of IRS 1C and 1D of 188 m are being used for detailed assessment of agricultural drought at district and sub district level in four states namely, Andhra Pradesh, Karnataka, Haryana and Maharashtra. States reports depicting the drought conditions are also published by the department for few of the past events (2002 drought). These reports can be used in identifying the districts which are chronically prone to drought like situation and also helpful in analysing the trend (once in 10years/ 5 years/ 2 years etc.). A comprehensive Indian **Forest Fire Response and Assessment System (INFFRAS)** is invoked under DSC activities of NRSC, which integrates multi–sensor satellite data and ground data through spatially and temporally explicit GIS analysis frame work. The INFFRAS is designed to provide services on Fire alerts i.e. value–added daily daytime TERRA/AQUA MODIS fire locations and DMSP– OLS derived daily night–time fire locations, Fire progression: Progression of fires using daily day and night fire location information given by MODIS/DMSP– OLS, burnt area expansion derived from temporal high resolution data sets and Burnt area assessment the mapping episodic fire events using moderate and high resolution optical data sets and preparation of forest fire mitigation plans. DOS has taken up the **Landslide Hazard Zonation Mapping** project in the year 1999 covering the major tourist / pilgrimage routes in the states of Uttarakhand and Himachal Pradesh. NRSC has carried out various studies on land use/land cover mapping, land cover change, etc.

NRSC has carried out land use/land cover mapping for agro–climatic zone planning covering the 442 districts in the country on 1:250,000 scale using two season (Kharif and Rabi cropping season) data for Planning Commission, Govt. of India. National Wastelands Inventory Project (NWIP) was undertaken by the National Remote Sensing Centre for the entire country on the Scale of 1:50,000 using Landsat TM and IRS LISS II/III data. This was the first time that such a gigantic project was undertaken to provide spatial and statistical information on the wastelands in the country so that appropriate land use planning could be undertaken for the restoration and development of such degraded lands. The study estimated about 63.85 M. ha (20.17%) as wastelands in the country in 13 different categories. Subsequently, these maps were updated under the Nation Wide Wasteland Updation Mission project during 2003–04. ISRO has also initialised the development of a **National Data Base for Emergency Management** for multi–hazard prone districts which will include base maps, hazard maps, resource maps and vulnerability indicators (census) at 1:50,000 scale and high resolution maps for selected urban areas. NRSC produces hazard maps and also post disaster data for assessing damages.

3.3 Disaster database and statistics

3.3.1 Indian databases

1. **Disastrous weather events**

   India Meteorological Department publishes a report on the impact of disaster in the country, The Disastrous Weather Events Report, since 1967.The summary of these natural hazards is based on the various reports of the India Meteorological Department and press information. The figures for estimated losses are as per the press reports and the areas affected by the adverse weather phenomena are depicted season wise on four maps for **12 hydro – meteorological disasters.** Date/period, city/area affected intensity, casualties and extent of damage due to snowfall, cold wave, heat wave, drought, squall, gale, dust storm, lightning, thunderstorm, hailstorm, floods and heavy rains, landslides and cyclonic storm are included in this report.

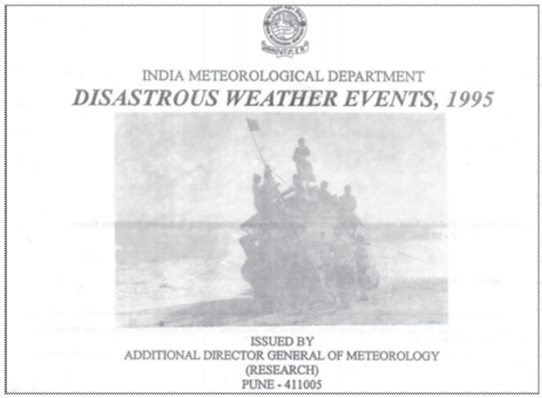


Figure A3.8: Coverpage of Disastrous Weather Events Report, IMD

2. **Daily Report on Rainfall & Flood** during the south–west monsoon (w.e.f. from 1st June to 31st September)

   State Government sends daily situation reports to the Control Room, Ministry of Home Affairs on the rainfall and flood condition during the entire duration of south west monsoon. This report includes the rainfall position (place and rainfall in cm), cause of floods, extent of damage, crops affected, houses damaged, damage to infrastructure, rescue and relief operation details. This reports are available online since 2004 http://ndmindia.nic.in/.

3. **State Government submit memorandum to the Ministry of Home Affairs** requesting for central assistance to meet the expenditure for providing immediate relief to victims of cyclone, drought, earthquake, fire, flood, tsunami, hailstorm, landslide, avalanche, cloud burst and pest attack (13th Finance Commission Report). These reports provide details of the causes, sectoral impacts, damages and losses and relief measures for selected major calamities. However these reports only cover major disasters and do not capture many small and medium scale events and disasters like heat and cold wave, erosion etc.

4. **Ministry of Agriculture (MoA)** was responsible to coordinate relief measures, on behalf of the Central Government, in the event of Natural Disasters and to also handle the subject of Natural Disaster Management generally till 2002. Presently MoA is responsible for the co–ordination of relief measures necessitated by drought, hailstorm and pest attacks and also matters relating to loss of human life due to drought. State–Wise details of districts declared drought and drought situation reports are maintained by drought management cell of MoA. Besides Directorate of Economics and Statistics of MoA publishes Agricultural Statistics of India (annual). A wide range of data related to land use, area under irrigation, crops, productivity, price and disaster related information is available at the website **http://lus.dacnet.nic.in/dt_lus.aspx**. Specific information on drought scenario and management strategies are available at **http://agricoop.nic.in/DroughtMgmt/Archives.htm.**

Figure A3.9: Webpage of Drought Management Cell, Ministry of Agriculture

5. **Integrated Disease Surveillance Project (IDSP)** was launched with World Bank assistance in November 2004 to detect and respond to disease outbreaks quickly. The project was extended for 2 years in March 2010. From April 2010 to March 2012, World Bank funds were available for the Central Surveillance Unit (CSU) at NCDC & 9 identified states (Uttarakhand, Rajasthan, Punjab, Maharashtra, Gujarat, Tamil Nadu, Karnataka, Andhra Pradesh and West Bengal) and the rest 26 states/UTs were funded from domestic budget. District wise incidences of epidemics and outbreaks are available on weekly basis in the report. **http://idsp.nic.in/.**

Figure A3.10: Website of IDSP Programme, Ministry of Health

6. CAIRS project has been developed by the Environment & Forest Informatics Division of NIC in close consultation with HazardousSubstances Management Division of the Ministry of Environment & Forests for capturing chemical accident data. Although the portal is fully functional, data is not captured/ updated for many states – **http://cairs.nic.in/.**

Figure A3.11: Website of Chemical Accident Information and Monitory System, MoEF

7. SDMA –State Disaster Management Authorities are developing mechanisms for regularly maintaining data on disaster events, damages and loss and the relief measures since 2005. Few states like Orissa, Uttarakhand, UP, Mizoram and Tamil Nadu developed historic disaster databases from various sources under the Indis data/ DesInventar project. Recent disaster reports are available at the website of BSDMA for Bihar (http://disastermgmt.bih.nic.in/), Disaster Management Department of UP rahat@up.nic.in, Karnataka KSDMA http://www.ksdma.co.in/DisasterInKarnataka.aspx and several other states.

3.3.2 International databases

Table A3.1: Criteria for including an event in the database adopted in various international databases

Indicator	CRED–EMDAT	IFRC	DESINENTA	EMA Database	SHELDUS	NOAA
Deaths	10 or more people	10 or more people	Any number	Any number	No criteria	One or more person
Injured	Not used as a criteria	Not used as a criteria	One or more person			One or more person
Affected	100 or people	100 or people	One or more person	One or more person	No criteria	One or more person
Property Loss			property loss/damag e to environmen	property loss	$50,000	One or more person
Emergency Declared	declaration of a state of emergency		No such criteria	No such criteria	No criteria	One or more person
Call for external assistance	call for international assistance		No such criteria	No such criteria	No criteria	One or more person

DesInventar: It is an inventory system, a methodology to register data about characteristics and effects of diverse types of disasters, with special interest in disasters that are invisible from global or national scales. DesInventar is a conceptual and methodological tool that captures disasters of any magnitude and about local, regional and national surrounding diversity. It is an inventory system, a methodology to register data about characteristics and effects of diverse types of disasters, with special interest in disasters that are invisible from global or national scales. The inventory allows watching accumulated data of these invisible disasters at a global or national scale. This tool has been widely adopted in Latin American countries. Customized version of DesInventar has been adopted by many Indian states (Orissa – Indis data / Tamil Nadu/ UP/ Mizoram/ Orissa), Nepal, Philippines, Indonesia, Srilanka and Maldives) **http://www.desinventar.net/.**

Figure A3.12: Homepage of DesInventar

3.3.3 CRED EM–DAT

The Emergency Disasters Data Base (EM–DAT) managed by by the Centre for Research on the Epidemiology of Disasters (CRED) at the Catholic University of Louvain, Belgium is a publicly accessible international database collecting information on natural and technological disasters. The database contains approximately 15,700 entries with an average of 700 new entries per year and covers the period from 1900 to the present. It is updated on a daily basis and made available to the public once a month after validation of the figures. Criteria for inclusion in the database are as follows: $\geqslant$ 10 people killed, and/or $\geqslant$ 100 people reported affected, and/or a declaration of a state of emergency, and/or a call for international assistance (4). Events are entered on a country–level basis and information collected includes location, date, number of people killed/injured/affected, number of homeless, and estimated damage costs. Sources include governments, UN agencies (UNEP, OCHA, WFP, and FAO), NGOs (IFRC), research institutions, insurance institutions (Lloyds) and press agencies, although priority is given to UN agencies. Amongst disaster databases, EM–DAT provides one of the most comprehensive and transparent explanations of the methodology employed. The database is searchable by country, disaster type, or timeframe. Due to the nature of the inclusion criteria, EM–DAT maintains a "global observation level and a national resolution level" (8). This makes some smaller scale disasters "invisible" and creates difficulties for agencies attempting to disaggregate disasters at a local or municipal level.The main objective of the database is to serve the purposes of humanitarian action at national and international levels. It is an initiative aimed to rationalize decision making for disaster preparedness, as well as providing an objective base for vulnerability assessment and priority setting. For example, it allows one to decide whether floods in a given country are more significant in terms of its human impact than earthquakes or whether a country is more vulnerable than another.EM–DAT contains essential core data on the occurrence and effects of over 12,800 mass disasters in the world from 1900 to present. The database is compiled from various sources, including UN agencies, non–governmental organisations, insurance companies, research institutes and press agencies. Since 1988 the WHO Collaborating Centre for Research on the Epidemiology of Disasters (CRED) has been maintaining an Emergency Events Database – EM–DAT. EM–DAT was created with the initial support of the WHO and the Belgian Government.

Figure A3.13: Homepage of EM–DAT, CRED

Source: www.emdat.be

3.3.4 Asian Disaster Reduction Centre (ADRC)

The Asian Disaster Reduction Centre (ADRC) maintains a database of disaster information for member countries, which provides information about natural disasters in Asia and Southeast Asia. Information is available in English and Japanese and the database is searchable by member country with entries beginning in 1998 and continuing to the present. Events are listed in 10 chronological orders and therefore easily searchable if an exact date of the event is known. It is not clear from the website what the criteria for inclusion are as there are few entries listed for each country. Disaster entries include the GLIDE number, location of event, date/duration, number of dead, and injured, evacuated, and material damages as reported by contributing sources. Sources include UN agencies (OCHA), Reuters and international news agencies (AFP, BBC, CNN), and NGOs (IFRC, Catholic Relief Services). Accessing disaster information can be a time consuming and laborious task. Not only is data scattered but frequently identification of the disaster can be confusing in countries with many disaster events. To address both of these issues, Asian Disaster Reduction Centre (ADRC) proposed a globally common Unique ID code for disasters. This idea was shared and promoted by the Centre for Research on the Epidemiologyof Disasters (CRED) of the University of Louvain in Brussels (Belgium), OCHA/ReliefWeb, OCHA/FSCC, ISDR, UNDP, WMO, IFRC, OFDA–USAID, FAO, La Red and the World Bank and was jointly launched as a new initiative "GLIDE".

http://www.glidenumber.net/glide/public/search/search.jsp

3.3.5 Australia: Emergency Management Australia Disasters Database

The EMA Disasters Database (Emergency Management Australia) sponsored by the Australian government is the primary resource for all national level natural and technological disasters. Available in the public domain the database has recorded events from 1622 to the present and includes events in which any or all of the following occurred "= 3 or more dead, 20 injured or ill, and/or significant damage to property, infrastructure, agriculture or environment, or disruption to essential services, commerce, industry, or trauma or dislocation of the community at an estimated total cost of A$10,000,000 or more". Detailed information on the duration of the event, number of people killed, injured, affected, homeless, and evacuated along with an accounting of economic loss including insured loss is available. Location is limited to area and region and few maps are available. The database is searchable by region, zone, date, or through an advanced search by disaster effect (people killed, cost range). Although there is no indication of a primary source, the database draws on information from government, the Insurance Council of Australia, emergency service agencies, research bodies, educational institutions, and press media.

http://www.ema.gov.au/ema/emaDisasters.nsf

3.3.6 St. Lucia Disaster Matrix

The government of St. Lucia through NEMO (National Emergency Management Organization) has developed the St. Lucia Disaster Matrix. Presented in spread sheet format, the matrix provides information on natural and technological disasters for St. Lucia. The database lists 55 entries spanning from 1780–2004 and includes information on the date of the event, the number killed, homeless, the cost of the event, and comments. Historical records of earthquakes prior to 1990 are included in the matrix however after 1990 earthquakes are recorded on separate spread sheets, which although including geophysical information (latitude/longitude, depth) do not provide any human, or economic loss data. Information is provided by government, UNESCO/PAHO, academic institutions, and media. It does not appear that the database has been updated since November 2004. There are no links to contact information or an explanation of the methodology available online – http://www.geocities.com/slunemo/disaster_matrix/history_index.html.

3.3.7 Canada: Canadian Disaster Database

The Canadian Disaster Database (CDD) is an on–going effort by the Office of Critical Infrastructure Protection and Emergency Preparedness (OCIPEP) to record natural, technological, and conflict–related disasters that directly affect Canadian citizens. Although mainly a national database, there are entries for international incidents in which Canadian citizens were involved. The database contains 700+ entries beginning in 1900 and going until the present day. The database is searchable by location (province), disaster type, or time–frame. Information provided includes a description of the event, location, number of dead, injured, evacuated, and economic losses. The database also provides references for the sources of information for loss estimates. Primary sources appear to be press media but information sources include government, provincial emergency management organizations, municipal government, insurance bureau of Canada, and NGOs along with occasional cross–referencing of EM–DAT for validation of findings. Inclusion criteria are made available on the website although more detailed methodology is not easily accessible – http://www.psepc–sppcc.gc.ca/res/em/cdd/search–en.asp.

3.3.8 United States: SHELDUS

The Spatial Hazard Event and Losses Database for the United States (SHELDUS), created and maintained by the University of South Carolina, is a publicly accessibly national level (United States not including Puerto Rico, Guam, US territories) natural disaster database. The database has recorded over 400,000 events since 1960 disaggregated down to the county level. Although "prior 1995 only events that generated a total loss of $50,000 in damage" were recorded, "post 1995 all events represented in the NCDC Storm Data with a specific dollar amount" is entered into the database. Sources of information include NCDC's Storm Data Publication, National Geophysical Data Centre (NGDC) Tsunami Event Database (www.ngdc.noaa.gov), and the Storm Prediction Centre. The database is searchable by date of event, hazard type, or state/county affected. Information provided includes the beginning date, the event type, location (including FIPS code for spatial information), and number of fatalities/injured, property damage and crop damage, cost estimates. It is possible to download the search results and the search metadata results. The database provides detailed geo–referencing and maps of event locations useful for county–level disaster mitigation.http://www.sheldus.org.

3.3.9 States: United States Storm and Hazard Database

The United States Storm and Hazard Database maintained by the US National Climatic Data Centre (NCDC) records national level publicly accessible natural disaster information. Records are maintained from 1950 to the present and are updated monthly. Information is gathered from NOAA Storm Data, NWS (which draws information from county/state/federal management offices, law enforcement, skywarn spotters, NWS damage assessment, press media, insurance industry and general public), and Storm Prediction Centre. The database is searchable by date of event, location, or event type although it is also possible to conduct a more detailed search by cost of damages, number of fatalities/injuries, and/or hail size/wind speed. Due to the enormity of the database, (there appear to be over 500,000 events entered and no explanation of inclusion criteria) searching for specific events requires a degree of precision. Information presented in each entry includes location (detailed to the zone), date/time, and type of event, magnitude, number of deaths /injured, property damage and crop damage. http://www.gesource.ac.uk/hazards/usastorms.html.

3.3.10 United States: National Hazard Statistics

NOAA National Weather Service through the Office of Climate, Water, and Weather Services provides a searchable database of national weather–related fatalities associated with a limited number of natural disasters. The website provides statistics on the following natural disasters; lightning, tornado, heat, cold, flood, wind, winter storm, tropical storm. Yearly summaries available from 1995–2004 for individual weather events include breakdowns of fatalities by state and by gender/age of deceased. US summaries and state summaries are available yearly from 1995–2004, include fatality information along with injured and an accounting of the cost of property damage and crop damage. Information is provided by NOAA's Storm Data report.

http://www.nws.noaa.gov/om/hazstats.shtml

3.3.11 Philippines

DSWD–DROMIC

The Department of Social Welfare and Development (DWSD) of the Government of the Philippines has developed a Disaster Response Operations Monitoring & Information Centre (DROMIC). The records begin in January 2006 and cover the past three months. There is no readily available documentation to indicate history of the initiative or current methodology and there does not appear be an archive of past records. Sources of information appear to be the 15 DSWD field offices along with national seismology and geophysical services, NGOs, and media. The information provided includes date and time of event, type of event, location (down to the level of house address), affected areas and populations, casualties, injured, a summary of the event and the assistance given, and a detailed cost of assistance. Events are chronologically entered so only searchable by date – http://disaster.dswd.gov.ph/.

Philippines: NDCC database

The National Disaster Coordinating Council (NDCC) through the Office of Civil Defence in the Philippines began compiling data for a national disaster database in 1990. Originally intended as a means of collecting information on destructive tropical cyclones, the database now collects data on natural and technological disasters. On the national level, the Office of Civil Defence provides data while on the local level, the agencies involved include the Offices of Social Welfare, local planning development offices, and local emergency management offices. The database is intended as being publicly available however at the time of this publication it was unable to be accessed via the web – http://www.ndcc.gov.ph.

3.3.12 Technological disasters: UNEP/APELL

The United Nations Environmental Program (UNEP) provides a publicly accessible international technological disasters database prepared by APELL (Awareness and Preparation for Emergencies on a Local Level). Data are presented on 316 chemical events occurring between the years of 1970–1998 and include information on the origin of the accident, date/year, location, products involved, and the number of people dead/injured/evacuated. Inclusion criteria include any or all of the following; ⩾ 25 deaths, ⩾ 125 injured or more, ⩾ 10,000 evacuated or ⩾ 10,000 people deprived of water. The database is searchable by any of the variables listed above, making information easy to access, though information provided is rudimentary and no economic loss information is recorded. Sources of information include OECD, MHIDAS, BARPI, Sigma, press media, and industry to name a few but no indications as to primary source are given making the origin and the validity of the information presented questionable.

3.3.13 Global Disaster Alert and Coordination System – GDAC

GDAC daily newsletter provides an overview of the natural disasters that happened in the last 24 hours and response measures for on–going disasters. GDAC reports currently cover earthquakes, tsunamis, tropical cyclones, volcanic eruptions and floods. The colour coding (red, orange and green) is related to the estimated humanitarian impact of the event. Up–to–date media coverage, latest maps and Relief–Web content related to these disasters are also available in the GDAC Site – www.gdacs.org.

3.4 Data sets on development and vulnerability indicators

3.4.1 Census of India

Census of India is having demographic and housing details since 1881. Statistics on demographic & socio – economic characteristics area, administrative divisions, population, institutional & houseless

Figure A3.14: Website of Census of India population, population density, rural–urban distribution, sex ratio, broad age groups, number of literates & literacy rate, work participation rate, religious composition, scheduled castes & scheduled tribes population are available at national, state, district, sub–district, towns, village levels. Census data is the major source of information for vulnerability and capacity indicators (http://censusindia.gov.in/).

3.4.2 Central Statistical Organization and National Sample Survey organization (NSSO)

The Central Statistical Organisation is responsible for coordination of statistical activities in the country, and evolving and maintaining statistical standards. Its activities include national income accounting; conduct of annual survey of industries, economic censuses and its follow up surveys, compilation of Index of industrial production, as well as consumer price indices for urban non–manual employees, human development statistics, gender statistics, imparting training in official statistics, five year plan work relating to development of statistics in the states and union territories; dissemination of statistical information, work relating to trade, energy, construction, and environment statistics, revision of National Industrial Classification, etc. National Sample Survey Organisation (NSSO) conducts nationwide sample surveys on various socio–economic issues in successive rounds, each round covering subjects of current interest in a specific survey period. NSSO has been conducting multi–subject integrated sample surveys all over the country in the form of successive rounds relating to various aspects of social, economic, demographic, industrial and agricultural statistics.Statistical hand books published by the State Department Statistics and Environment is having data sets/ statistics on area & population, state domestic product estimates, climate & rainfall, vital statistics, agriculture & livestock, industry, energy & electricity, trade & finance, transport & communication, education, and health, labour & employment, police, crimes & accidents, excise statistics, water supply & sewage disposal, economic census, social & scheduled caste/scheduled tribe welfare and other miscellaneous statistics.

The Directorate of Economics and Statistics functions as the statistical authority as well as the nodal agency for collection, compilation, analysis, interpretation and dissemination of statistical information relating to all sectors of the economy of the states. The Directorate of Economics &Statistics, is the nodal agency for collection, compilation, analysis of statistics relating to various sectors of economy. The main function of the directorate is to build up a firm and broad database, which is essential for the formulation of sound policies and planning. In short, the Directorate of Economics and Statistics acts as a storehouse of statistical information in respect of the state and as a channel for feeding the statistical information to user organisations.

Besides, the district level Statistical Hand Book with a number of statistical tables on a wide range of subject fields provides statistical information on the district's like area and population, education, vital statistics, employment, administration, climate and rainfall, housing and building activity, agriculture, rural water supply, irrigation, local bodies, animal husbandry, civil supplies, fisheries, road and transport, forests, communications, electricity, industries, recreation, factories, social welfare, tourism, environment, medical and health, disaster related information etc. developed for all districts of India.

3.4.3 Compendium of environmental statistics

The Central Statistical Organisation brought out eleven issues of the publication entitled "Compendium of Environment Statistics" for the years 1997, 1998, 1999, 2000, 2001,2002, 2003, 2006, 2007, 2008–09 presents available data relating to environment of the country. Although, the present coverage of information in the compendium may not be exhaustive with respect to entire domain of environment, it does however provide a glimpse of the present scenario of the environmental degradation, its causes and the reasons for concern. It provides the necessary base to bring out the magnitude of the problem. The compendium consists of seven chapters: Introduction to environment, its degradation through different sources and their impact on human health and the development of environment statistics in India. The remaining five chapters are on Biodiversity, Atmosphere, Land/soil, Water and Human Settlements. Besides, statistical tables depicting environment data, suitable graphs and charts have also been added in the publication – http://mospi.nic.in/.

The State Departments of Economics and Statistics in several states developed compendiums of environment statistics which capture wide range of information on state of environment including natural resources, quality of water, air and land, hazard related and disaster related information.

3.4.4 Census Info/ DevInfo and emergency info

DevInfo is a database of social and development indicators, developed by the Ministry of Statistics and Programme Implementation in collaboration with the UN System in India. DevInfo is a powerful database system which monitors progress towards the Millennium Development Goals. It generates tables, graphs and maps for reports and presentations.Emergency Info is a powerful decision support system, based on the DevInfo database technology that helps people to respond better in emergency situations. It combines the advanced data access and presentation features of DevInfo with new data capture technologies. Emergency info helps to bridge information gaps within the first 72 hours of an emergency and provide support for rapid data collection, situation assessment, standard monitoring reports and disaster preparedness. For details visit http://www.devinfoindia.org/ and http://www.devinfo.info/emergencyinfo/.

Figure A3.15: Website of DevInfo India

3.4.5 Planning Commission / State Planning Departments

National human development reports by the Planning Commission and State Human Development Reports are having statistical appendix showing the state of human development, indicators on economic attainment, amenities, educational attainment, and health attainment & demography. Other attainments and some indicators on governance are available in the national HDR. Twenty two states of India so far released State Human Development Reports (SDR) details available at http://www.undp.org.in/state_human_development_reports. SDR of Orissa has a chapter on disaster vulnerability as well. Similarly HDR Assam also has the vulnerability profile. District Human Development Reports are available for few districts like Malda and Bankura. http://www.planningcommission.nic.in/http://www.planningcommission.nic.in/data/datatable/dt_pophsd.pdf.

3.4.6 National Informatics Centre

National Informatics Centre (NIC) under the Department of Information Technology of the Government of Indiais actively working for the past three decades in the area of Information and Communication Technology (ICT) applications in the government sector. NIC plays a pivotal role in collecting, compiling and disseminating information at national, state and district levels. NIC has already developed state and district profiles for all the states of India. Under the National Spatial Database Infrastructure project demography and amenity mapping up to village level has carried out by NIC GIS Division and is available in the public domain.

3.4.7 Environmental Information System (ENVIS)

ENVIS is a decentralised system with a network developed by Ministry of Environment and Forests in 1982, for ensuring integration of national efforts in environmental information collection, collation, storage, retrieval and dissemination to all concerned. Presently the ENVIS network consists of focal point at the Ministry of Environment and Forest and ENVIS centres setup in different organisations/establishments in the country in selected areas of environment. These centres have been set up in the areas of pollution control, toxic chemicals, central and offshore ecology, environmentally sound and appropriate technology, bio–degradation of wastes and environment management, etc. ENVIS is also having an Indian State Level Basic Environmental Information Database.

3.4.8 Other sources

Other sources of information are various central ministries i.e. Ministry Agriculture and Cooperation, Health and family Welfare, Home Affairs, Ministry of Mines, Earth Sciences, Department of Science and Technology. Other departments like National soil survey and land use board, State department of Environment, Land use, Town and Country Planning, Urban Development, Rural development, Department of Agriculture, State and district units of National Informatics Centre, Regional Remote Sensing Application Centres, State Remote Sensing Application Centres (List available under web references at the end of this module).

3.5 Resource inventory

3.5.1 India Disaster Resource Network

India Disaster Resource Network is an online inventory developed by Ministry of Home Affairs with the technical support of United Nations Development Programme. The portal has been designed to serve as a decision making tool for the government administrators and crisis managers to coordinate effective emergency response operations in the shortest possible time.

Figure A3.16: Website of India Disaster Resource Network

IDRN is having nation–wide electronic inventory of resources that enlists equipment and human resources, collated from district, state and national level Government line departments and agencies. The inventory also has informative details of NGOs and private sector including around 5000 corporate members registered with the Confederation of Indian Industry and 33000 builders, contractors and construction companies registered with Builder's Association of India and other public sector undertaking. For details visit www.idrn.gov.in.

3.5.2 State Disaster Resource Network – Gujarat

Figure A3.17: Website of State Disaster Resource Network

State Disaster Resource Network [SDRN] is a state owned centralised database system, which collects, stores and processes the detailed infor–mation up to village level and generates query based reports on a) hazards profile, b) vulnerable elements at risk, c) resources available, d) disaster history and e) emergency contacts to strengthen the emergency response preparedness and disaster risk reduction in the State of Gujarat. For details visit http://sdrn.gujarat.gov.in/.

3.5.3 Other Sources

Data of resources are also available in district and state level disaster management plans and village level DM plans (in tabular format).

3.6 Recent initiatives

About India's data project and DesInventar Database

For developing systematic geo–referenced historical disaster databases with all the natural and manmade disasters of different magnitude, intensity and frequencies at the district and sub–district levels, India's data project has been initiated by BCPR– UNDP in Orissa. About 10,000 disaster event report of 30 years period for the state of Orissa has been collected and spatial and temporal distribution of various disasters and its linkage with vulnerability/ development indicators has been analysed. Similar exercise of compiling historical database and developing and disaster statistical system has been initialized in Tamil Nadu under the Regional Tsunami Recovery Programme and for the state of Uttar Pradesh, Uttaranchal and Delhi under Disaster Risk Management Programme. However in the absence of well–established institutional framework and mechanisms the database is not updated and most of these initiatives remind as pilot interventions only.

About CSO – NIDM national disaster statistical system

At present data on different aspects of disasters are collected by various agencies, but mostly these remain confined in newspapers reports or official files and sometimes periodical bulletins and reports are published by some agencies, but there is no system to collect, compile, validate and publish such data in one place on a regular basis for the use of policy makers, analysts, disaster managers and other users. Therefore a need has been felt for the development of a comprehensive national database on disasters and institutionalizing a national disaster statistical system in India which would facilitate formulation of area specific disaster risk profile, assessment of long term impacts of disasters, development of policies, strategies and frameworks, preparation of proper planning for disaster preparedness and allocation of adequate funds for the prevention and mitigation of disasters etc.Brainstorming workshop on the development of a National Disaster Statistical System was organized by the Central Statistical Organisation (CSO), Ministry of Statistics and Programme Implementation (MoSPI) in collaboration with the National Institute of Disaster Management (NIDM), Ministry of Home Affairs on 27 April, 2007. Two committees, one on hazard related statistics and other on disaster statistics has been set up at national level with members from various nodal ministries and departments.

More recently Government of India initiated steps towards integration of disaster data with the National Statistical System. Formats for collection and compilation disaster related data were developed by CSO in consultation with NIDM and MHA. These formats were sent to State Directorate of Economics and Statistics for facilitation of data collection from states. Formats for data collection from states were finalised by CSO and send to states in the month of May 2012. These formats are systematic in nature and capture data from district level on disaster event, cause, impact, relief and rehabilitation.

India Disaster Report – NIDM

National Institute of Disaster Management published first India disaster report for the year2011 in June 2012, in this report, detailed information about the causes, impact, response and relief and lessons learned from major disasters for which the stategovernments had requisitioned the Central Government for supplementary support. In addition, few disasters like the stampede in Sabarimala pilgrimage and re breakout in hospital in Kolkata have also been discussed in detail due to the large number of lives lost in these events. A list of disaster events in which ten or more human lives were lost is annexed at the end of the report. The list has been primarily drawn from the daily disaster update compiled by the National Institute of Disaster Management for the reported disasters in 2011.

http://nidm.gov.in/PDF/India%20Disaster%20Report%202011.pdf.

3.7 Gaps and challenges

Although the existing databases have tremendous potential in understanding the hazards and vulnerability and climate–change patterns and its projected impacts especially at large scale, the following gaps and challenges still exists. In India and in most of the countries, information about past disasters has not been documented systematically. Some information would be available in various government departments dealing with relief and rehabilitation work. However, this would be restricted to major disasters, and information about small events such as domestic fire etc. is not documented. Hazard specific data are not collated to be used in integrated approach and are available at different temporal and spatial scales. Most of the tools and datasets are limited to single hazard type and linkages of the causative factors of various disasters and also their impacts in the context of climate–change are not available.

Knowledge base about the existing datasets such as information on hazard, vulnerability and risk are not available at unified source. Information on various hazards and their triggering factors are available to some extent but not integrated. There is no unified databank on all disasters and their temporal and spatial distribution, causes, impacts on different sectors in short–term and long–term scales, as well as various relief and rehabilitation measures taken is available. There are well established systems at various levels for collecting data, viz. socio–economic indicators, demographic, land records, health information, etc. but there is no well–established system for timely collection, compilation and analysis of disaster data

4

BIBLIOGRAPHY MODULE A

BMTPC (1997).“Vulnerability Atlas of India”, Building Materials and technology Promotion Council.Ministry of Urban Development, Govt. of India, New Delhi.

BMTPC (2006).“Vulnerability Atlas of India”, Building Materials and technology Promotion Council.Ministry of Urban Development, Govt. of India, New Delhi.

Briceño S. (2008). Disaster Risk Reduction and Climate Change Adaptation an International Perspective, From the past natural disasters to the challenge of climate change in Europe – Paris 26–28 November 2008, United Nations International Strategy for Disaster Reduction.

CDERA (2002). Report of the Brainstorming Workshop on Adaptation to Climate Change in Caribbean Disaster Risk Management, Caribbean Disaster Emergency Response Agency, June 6 – 7, 2002.

Chandra P. and Gupta S.K. (Eds.) (2005). Inventory of landslides of Northwest Himalaya: With available data from Eastern Himalaya. Kolkata: Geological Survey of India.

CWC (2009). National Register of Large Dams. New Delhi: Central Water Commission, Govt. of India. Retrieved from www.cwc.nic.in/main/downloads/National%20Register%20of%20Large%20–Dams%202009.pdf

CWC (2010). Water and Related Statistics.New Delhi: Central Water Commission. Retrieved from http://www.cwc.nic.in/ISO_DATA_Bank/W&RelatedStatatics_2010.pdf

CWC (2012). Integrated Hydrological Data Book.New Delhi: Central Water Commission. Retrieved from http://www.cwc.nic.in/main/downloads/Combined%20Final_HDD_09042012.pdf

CWC and NIH (2008). Preliminary Consolidated Report on Effect of Climate Change.Central Water Commission, New Delhi and National Institute of Hydrology, Roorkee. Retrieved from http://cwc.gov.in/main/downloads/Preliminary_Report_final.pdf

Creel L. (2003). Ripple Effects: Population and Coastal Regions, Making the Link, Population Reference Bureau, Washington, DC.

Dasgupta S. et al (2000). Seismo– Tectonic Atlas of Indian and its Environs, Geological Survey of India, Calcutta.

Dilley M., R.S. Chen, U. Deichmann, A.L. Lerner–Lam, M. Arnold, J. Agwe, P. Buys, O. Kjekstad, B. Lyon, and G. Yetman (2005). Natural Disaster Hotspots: A Global Risk Analysis: A Synthesis Report, The World Bank, Hazard Management Unit, Washington DC.

Douglas J. (2007). Physical vulnerability modelling in Natural Hazard Risk Assessment. Nat. Hazards Earth Syst. Sci., 7: 283–288

Doukakis E. (2005). Coastal vulnerability and risk parameters.EuropeanWater. 11/12: 3–7.

ESD (2004). Multi–Hazards Urban Risk Assessment with Dynamic Spatial Information for Tokyo.Environment & Sustainable Development Programme, United National University & University of Tokyo (http://www.unu.edu/esd/projects/hazardrisk.htm).

FAO (2007). Adaptation to climate change in agriculture, forestry and fisheries: Perspective,framework andpriorities. Report of Interdepartmental working group on climate–change, Food and Agriculture Organization of UN.Rome, Italy.

Finance Commission (2009). Disaster Relief. In: Report of the Thirteenth Finance Commission (2010–2015). New Delhi: 13th Finance Commission.

Geological Survey of India (1999). Inventory of the Himalayan Glacier.GSI Special Publication No.34.

Geological Survey of India (2009). An Inventory of Major Landslides in Sikkim – Darjeeling Himalaya.GSI Special Publication No. 92.

Gupta A.K., S.S. Nair and V.K. Sehgal (2009). Hydro–meteorological disasters and climate change: conceptual issues and data needs for integrating adaptation into environment and development framework. Earth Science India, 2(2): 117–132.

HPC (2001).The Report of High Powered Committee on Disaster Management. National Centre for Disaster Management, Indian Institute of Public Administration, New Delhi

ISDR (2002). Living with Risk, A Global review of Disaster Reduction Initiatives, prepared by International Strategy for Disaster Reduction, World Meteorological Organisation and Asian Disaster Reduction Centre

IMD (1967 to 2005) India. Disastrous Weather Events Report, Annual Publication of India, Meteorology Department, Pune.

NDMA (2009). National Policy on Disaster Management. National Disaster Management Authority, New Delhi. Retrieved from www.ndma.gov.in

NIDM (2012). India Disaster Report, 2011.http://nidm.gov.in/books

Shetye, S.R., A.D. Gouveia, and M.C. Pathak (1990). Vulnerability of the Indian coastal region to damage from sea level rise.Current Science, 59(3): 152–156.

Tompkins H. (2002). Climate change and extreme weather events: Is there a connection? Cicerone 3/2002. CICERO from the internship program International Institute of Sustainable Development (IISD)

UNDP–BCPR (2004). A global report on reducing disaster risk: A challenge for development. United Nations Development Programme, Bureau for Crisis Prevention and Recovery, New York.

UN–IDNDR (1994). Yokohama Strategy and Plan of Action for a Safer World: guidelines for natural disaster prevention, preparedness and mitigation. World Conference on Natural Disaster Reduction.Yokohama, Japan. Retrieved from http://www.preventionweb.net/files/8241_doc6841 contenido1.pdf

UN–ISDR (2005). Hyogo Framework for Action 2005–2015: Building the Resilience of Nations and Communities to Disasters.Extract from the final report of the World Conference on Disaster Reduction (A/CONF.206/6). Retrieved from http://www.unisdr.org/we/inform/publications/1037

UN–ISDR (2002). Living with Risk.A Global Review of Disaster Reduction Initiatives.Geneva, July 2002.

United Nations Environment Programme (2008). Data needs for addressing Climate Change – UNEP's perspective. UNEP/DEWA scoping paper; UNSD/Statistics Norway Conference on Climate Change & Official Statistics, Oslo, 14–16 April 2008.

Wisner B. et al. (2004). At Risk: Natural hazards, people's vulnerability and disasters. London, Routledge.

Module B

Introduction to statistical terminologies

Contents

Module B

1	**Lerning Unit B–1: Introduction**	65
2	**Learning Unit B–2: The scenarios**	67
	2.1 What is this unit about?	67
	2.2 The scenario – flood	68
3	**Learning Unit B–3: Scenario of Bihar**	71
	3.1 What is this unit about?	71
	3.2 A guided exercise: Look at some historical data from Bihar	74
	3.3 Example of rainfall figures from Bihar State	75
4	**Learning Unit B–4: Measures of central tendency and spread**	77
	4.1 What is this unit about?	77
	4.2 Distribution of data	77
	4.3 Measures of central tendency and spread	79
	4.4 Average or mean and variable scales	80
	4.5 Median, quartiles and mode	83
	4.6 Measures of spread: range and standard deviation	84
	4.7 Other measures of spread	85
	4.8 Rainfall figures from Bihar State	87
	4.9 Guided exercise (1)	89

5	**Learning Unit B–5: More on measures of dispersion**	91
	5.1 What is this unit about?	91
	5.2 Examples of measures of dispersion – rainfall figures from the Bihar State	92
	5.3 Guided exercise (2)	94
	5.4 How to use measures of dispersion for the analysis of your data?	98
	5.5 Resume of the examples of rainfall figures from Bihar State using statistics	100
	5.5.1 Results and preliminary analysis based on measures of dispersion	100
	5.5.2 Independent exercise of rainfall figures from Bihar State using statistics	101
6	**Self–Assessment for Module B**	103
7	**Bibliography Module B**	109

1

LEARNING UNIT B–1: INTRODUCTION

The purpose of the module is to deepen and broaden existing knowledge of statistical tools and terminology in the context of Disaster Risk Management, especially in India. The following pre–test is for orientation of the participants and the facilitators. The facilitator will use the results of the pre–test to adapt the course contents, especially the interactive and communication parts according to the knowledge of the participants.

In the introduction a brief description of the reason and the methods to collect data and to improve knowledge and decision making based on information is given. Of course, collecting data is not without reason and the purpose here is to relate data collection and the use of information to the goals of Disaster Risk Management. In order to better understand their uses and limitations and to enhance ability for designing perspectives of improvement and further development we ask:

Key questions which come to our mind are

- Why do we engage in the use of 'Statistics for Disaster Risk Management'?
- What can and – equally important – cannot be done with the use of 'Statistics for Disaster Risk Management'?
- What are the basic concepts, terminology and tools for the use of 'Statistics for Disaster Risk Management'?

Practice over theory

In the first learning unit introduction to databases, sources and applications are discussed.

Key objective of this unit are

- Understand and compare measures and techniques.
- Discuss outcomes and interpret results.

2

LEARNING UNIT B–2: THE SCENARIOS

2.1 What is this unit about?

In this unit we will describe the scenario in national terms and, in general terms, the use of statistics, the reason and the methods to collect data and to improve knowledge and decision making based on information.

We will select two typical types of scenarios: Flood and agricultural drought. Why? Look at the Figure B2.1 and the number of people affected by disasters in India:

Affected People

Disaster	Date	Affected	(no. of people)
Drought	1987	300,000,000	
Drought	2002	300,000,000	
Flood	1993	128,000,000	
Drought	1982	100,000,000	
Drought	2000	50,000,000	
Flood	2002	42,000,000	
Flood	1982	33,500,000	
Flood	2004	33,000,000	
Flood	1995	32,704,000	
Flood	1980	30,000,023	

Figure B2.1: People affected by disasters in India

Source: PreventionWeb

We clearly see that these two disaster types might not have caused the highest figures in human casualties or damages (we will come to this in detail in Module C) but a maximum number of people are affected almost every year by these two disasters.

Expected results:

The participants will understand why we choose this scenario and why they use fact based information for the benefit of Disaster Risk Management.

Enable the participants to interpret available rainfall data for assessing the hazard scenario.

Identify explicit actions for a Disaster Risk Management by participants while completing assignments.

In order to get more precise information, we have to look at a smaller area than the Indian sub–continent. We selected the State of Bihar. It could have been any other State, but Bihar has enough characteristics to qualify for the setting of a scenario (we will see this later in detail). In the areas of Disaster Risk Management we will deal with some exemplary districts of the whole State of Bihar and these two possible types of disasters.

It seems reasonable to select the rainfall data which are closely related to flood and drought and follow a natural annual cycle. Rainfall is irregular in Bihar State. In many years severe floods have affected human lives and assets but agricultural drought is a severe threat to human lives and economic well–being as well. Finding, reading and interpreting the rainfall data will be the purpose of this module. We will also talk about statistical terminology but strictly in the context which relates to Disaster Risk Management.

2.2 The scenario – flood

Why using data?

Disaster Risk Management is almost impossible without proper, timely and area related data. Decision making based on guesswork is always difficult and beset with inaccuracy. When it comes to disasters, decision making that ignores pertinent and available information is downright careless and unaccountable.

Let us select example of floods and droughts in Bihar. Since both floods and droughts are related closely to rainfall we will look at and analyse rainfall data for the State of Bihar. For the sake of simplicity the relationship of rainfall and flood/drought for Bihar will be looked at and the possible DRM actions discussed/proposed in this course will be related to these data only. The emphasis of this course is not the DRM activities because you as an expert will already and usually know best what to do. This course focuses on reading/collecting and analysing the right data to improve disaster mitigation, preparedness and response.

What can data tell us?

While investigating on data necessary to improve on Disaster Risk Management you will have to find answers to the basic and most pertinent questions, such as:

- Which area are we dealing with?
- Can we limit our search due to commonplace knowledge (no tsunami on a mountain)?
- What were the specific conditions in the past?
- At what time / under which conditions are the risk of a disaster high?
- At what time / under which conditions are the risk of a disaster low?
- Where are the possible damages and losses located?

If we find the data to help us to answer these questions, this would already be a of great use in decision making.

What data cannot tell us?

While depending on reliable information, many influences, constraints, conditions are beyond data capture and analysis, such as

- Statistics are general pieces of information that apply to an entire 'population' of people wherefrom the statistics were calculated. This could be tens, hundreds or thousands of people. Statistics can't tell you exactly what will happen to you personally. However, statistics may be able to tell you what the chances are of something like a disaster happening to you.
- If 65% of people with your type of illness responded to a treatment, then there is a two out of three chance that you will too. But no one can say definitely whether your illness will respond to that treatment.
- What will be the rainfall conditions during next year? (long–time weather forecast is very unreliable)
- What will be your personal risk of being struck by a disaster? (if prediction of a disaster for a group of people are high, you are very likely to be affected if you belong to this group)
- Statistics always talks about a population and not about individuals
- Statistics can't tell you about your future individual damages and losses

So to lie with statistics is done often but most of the time for two reasons:

If the author of existing statistics ignored and neglected some of the pre–conditions (like reliable sampling etc.) and rules of producing reliable statistics (like transparent methodology etc.)

- because he/she did not know better
- because he/she wanted to deceive the audience

There might be other reasons as well and we will discuss them later in module C: "Understand causes and effects of disasters as described by statistics" in detail.

Where will Disaster Risk Management come in?

The participants of our course should have at least basic computer knowledge especially in EXCEL and some understanding and/or curiosity and/or professional interest in the use of statistics for Disaster Risk Management.

Your action to make use of this course will depend on many things, as your working area will probably differ greatly as your working situation will probably differ very much.

It doesn't matter whether you're a journalist averting the public about a flood or whether you are the DRM Manager preparing for a future drought, you will always have to look at the related rainfall figures. This is what we will do in more detail in the following units. We will also propose some steps in order to take further action. If you have questions on the subject, you may find some answers to your questions in this course. As you follow the route which we have laid out for your understanding for the relationship between statistics and an application will improve your existing conceptual and practical knowledge shall be enriched.

It is believed that this course will improve your understanding of the use of statistics for DRM and of the kind of calculation and analysis that can be done in that framework. However, it will not replace the practical experience needed that can only be achieved by a professional involvement at a substantial level. In the Knowledge check, embedded in this course, you will have the opportunity to solve individually or in group mock exercises/simulations.

3

LEARNING UNIT B–3: SCENARIO OF BIHAR

3.1 What is this unit about?

In this unit we will describe the scenario in regional terms, describe the setting in Bihar, the physical and administrative structure and look at the available rainfall statistics. We will further identify the selected districts. Whereas these districts are somehow representative for the rainfall situation in Bihar, their selection had to be somewhat arbitrary because the working interest of the participants might be in another state in other districts.

Next we introduce the EXCEL template used later in the next unit for the "Guided Exercise" in this module. Participants will understand where to access rainfall data for selected districts in Bihar and how the rainfall data are displayed in the EXCEL spread sheet.

The "Guided Exercise" does not start here but the EXCEL template is introduced.

About the software used

The "Guided Exercises" use a version of EXCEL because its use is widespread and basic statistical functions are embedded. Other software could equally be used without any detriment especially Open Office, the open software tool (openoffice.org). For the statistics application the use of a free of charge 30–day test version statistical tool (StatistiXL) is advocated because of its integration into EXCEL and the superior visibility and clarity of results. This tool is integrated into EXCEL and rather easy to use.

Specialized statistics packages like SPSS, STATA (to recommend just two) will do the same job but with a much higher learning time requirements. All analysis of the "Guided Exercises" can be done without this extra software with on–board tools of EXCEL or OpenOffice but with some restrictions on the clarity of the results. Alternative functions in EXCEL / OpenOffice will be given on the exercises.

Expected results

The participants will understand how we chose and why we chose this scenario in Bihar and why they use fact based information for the benefit of Disaster Risk Management. The EXCEL Template is introduced and participants can read the file and understand its content.

Where is our scenario?

Our examples are selected from the State of Bihar:

Figure B3.1 : Bihar administrative districts in 2011

Source: http://censusindia.gov.in/2011census/maps/administrative_maps/BIHAR.pdf (last access 23.8.2013)

Few district data selected as examples. As one notices, rivers passes through 3 of the selected districts.

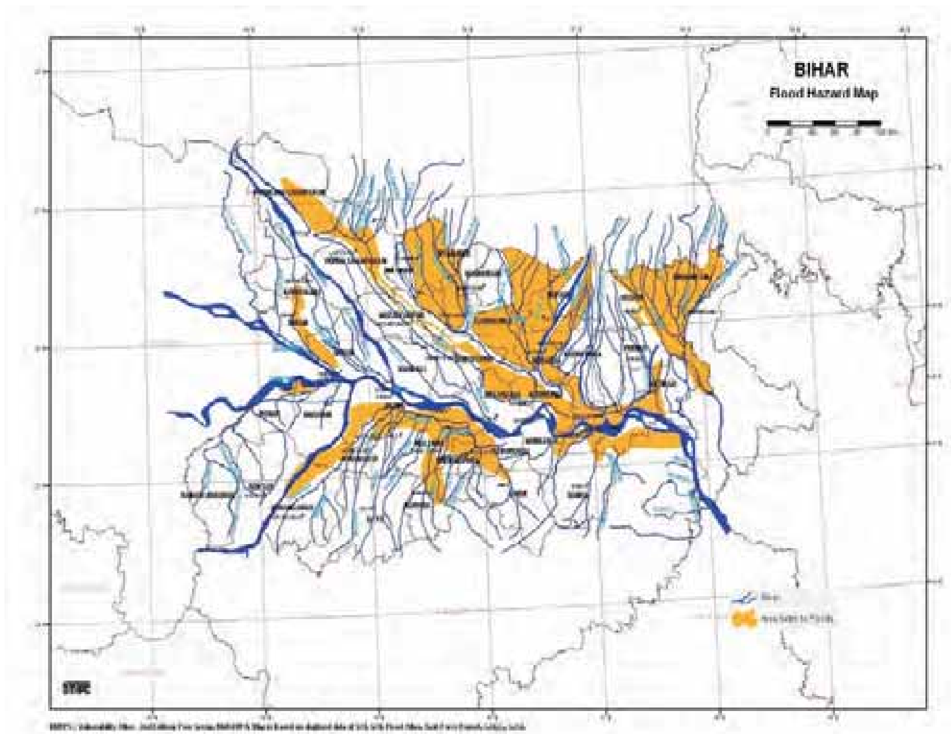


Figure B3.2: Bihar flood hazard map

Source: www.bmtpc.org (last access 23.8.2013)

What do we want to look at?

Our first interest will be to look at rainfall data. Where to get them from?

Best access through India **Meteorological Department** http://www.imd.gov.in/ (last access 23.8.2013).

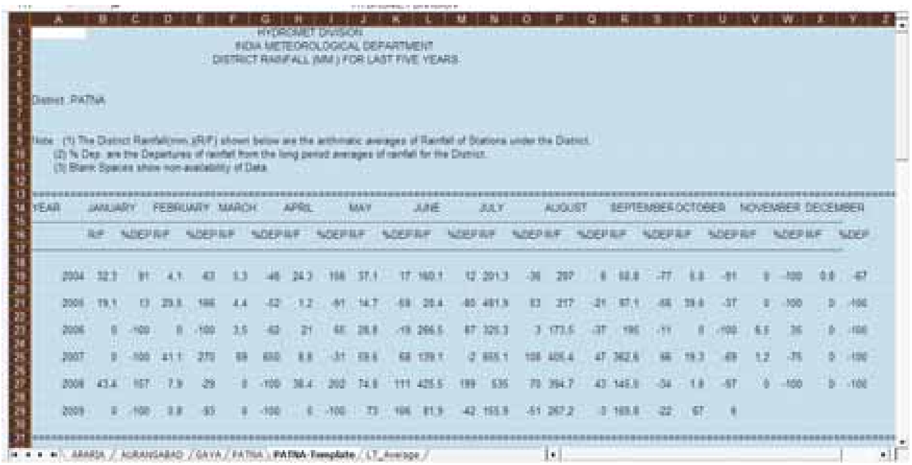
Figure B3.3: Website of the India Meteorological Department

3.2 A guided exercise: Look at some historical data from Bihar

Go to <Last 5–Years District wise Rainfall>

(http://www.imd.gov.in/section/hydro/distrainfall/districtrain.html) and select the appropriate districts.

Rainfall data will appear. The data is already been copied to the **EXCEL Template** for few districts. In the subsequent exercise you will do the same for another selected district. How this is done will be made clear in the next units and the sections “Guided Exercises”.



HYDROMET DIVISION
INDIA METEOROLOGICAL DEPARTMENT
DISTRICT RAINFALL (MM.) FOR LAST FIVE YEARS

District : PATNA

Note : (1) The District Rainfall(mm.)(R/F) shown below are the arithmatic averages of Rainfall of Stations under the District.
(2) % Dep. are the Departures of rainfall from the long period averages of rainfall for the District.
(3) Blank Spaces show non-availability of Data.

YEAR	JANUARY		FEBRUARY		MARCH		APRIL		MAY		JUNE		JULY		AUGUST		SEPTEMBER		OCTOBER		NOVEMBER		DECEMBER	
	R/F	%DEP	R/F	%DEP	R/F	%DEP	R/F	%DEP	R/F	%DEP	R/F	%DEP	R/F	%DEP	R/F	%DEP	R/F	%DEP	R/F	%DEP	R/F	%DEP	R/F	%DEP
2004	32.3	91	4.1	43	5.3	-46	24.3	158	37.1	17	160.1	12	201.3	-36	297	6	58.8	-77	8.8	-91	0	-100	0.8	-67
2005	19.1	13	29.8	166	4.4	-52	1.2	-91	14.7	-59	28.4	-80	491.9	53	317	-21	87.1	-66	39.6	-37	0	-100	0	-100
2006	0	-100	0	-100	3.5	-62	21	68	28.8	-19	266.5	87	325.3	3	173.5	-37	190	-11	0	-100	6.5	36	0	-100
2007	0	-100	41.1	270	59	650	8.8	-31	59.6	68	139.1	-2	855.1	108	405.4	47	362.6	66	19.3	-69	1.2	-75	0	-100
2008	43.4	157	7.9	-29	0	-100	38.4	202	74.8	111	425.5	199	535	70	394.7	43	145.0	-34	1.8	-97	0	-100	0	-100
2009	0	-100	3.8	-83	0	-100	0	-100	73	166	81.9	-42	153.9	-51	267.2	-3	169.8	-22	67	6				

ARARIA / AURANGABAD / GAYA / PATNA / PATNA-Template / LT_Average

Figure B3.4: Bihar rainfall data from the India Meteorological Department website

Source: http://www.imd.gov.in/section/hydro/distrainfall/webrain/bihar/patna.txt

3.3 Example of rainfall figures from Bihar State

What do we see?

- The PATNA–Template sheet contains the sample data for the PATNA District
- The lines represent the years
- The columns are displayed with monthly rainfalls (R/F) in mm and
- The departures of rainfall from the long period averages of rainfall for the district (%DEP) in percentages.
- The other sheets display the rainfall figures for the four selected districts: ARARIA, AURANGABAD, GAYA and PATNA.
- Another Sheet (LT_Average) displays rainfall normal (in mm) shown are the long period (1941–90) averages of rainfall in the State of Bihar for each month of the year and the seasonal subdivisions:

- Winter (JAN–FEB), Pre–Monsoon (MAR–MAY), Monsoon (JUN–SEPT), Post–Monsoon (OCT–DEC) and
- ANNUAL

What do we want to find out: Rainfall patterns (highs, lows, and dispersion of rainfall) for the months? The answers will be given in the next unit in the “Guided Exercises”.

4

LEARNING UNIT B–4: MEASURES OF CENTRAL TENDENCY AND SPREAD

4.1 What is this unit about?

In this unit we will explain the elementary statistical terms we use to describe rainfall patterns. Some of these terms are rather common places like Average; others are rather exotic, like Kurtosis. We will apply these descriptive terms to the rainfall situation in Bihar in the last section of this unit, the first part of the "Guided Exercise".

The "Guided Exercise" will lead you step by step on a path to apply the statistical terminology on a selected district. It is assumed that once you have done this for the district in the "Guided Exercises" you can do it on any other district in India (where data are available)

The statistical terms are explained in the context of rainfall patterns for the Bihar state. It is assumed that they will best be understood in this practical context

Expected results

The participants will understand what the statistical terms mean and why they are relevant for analysing rainfall data. The first part of the "Guided Exercises" shows you how to apply the template functions for other districts. In fact it is assumed that participants understand to apply this procedure for any district chosen in India (where data are available).

4.2 Distribution of data

In statistics, a Probability Distribution identifies either the Probability of each value of a Random Variable or the probability of the value falling within a particular Interval. The probability distribution describes the range of possible values that a random variable can attain. It also describes the probability that the value of the random variable is within any (measurable) subset of that range.

That sounds very theoretical but it amounts (as far as rainfall is the variable) to the simple fact that rainfall ranges between 0 and a maximum value, varying from district to district. These values are usually displayed in tables or graphs. Every New Year and every new month will generate a new value for this variable (the rainfall). Value (mm of rainfall) and distribution (where and when (?) rainfall can be expected) are of fundamental importance when it comes to floods and droughts.

There is a spread or Variability in almost any value that can be measured in a population (e.g. rainfall, height of people, durability of a metal, etc.) as well as an expected value for each month. However expectation and reality differ and this chapter is about how to describe this.

Let's first distinguish between Population and Sample:

- The Population constitutes the whole data set. In the case of the population of India it is the totality of individuals in India. When it comes to rainfall in a district of Bihar the "population" would comprise of all possible observable rainfall(s) in that district. Noticing the difference, in the first case, this number is finite, meaning it is a real number between for instance 1 and 2 billion, in the second case this number is unlimited (mathematically: infinite) because we can assume that the rainfall will vary (however marginally) each time we measure it. As the population is something we will never be able to grasp in its entirety, since we will neither be able to interview every single person living in India nor will be able to collect the data on every possible rainfall/single raindrop.
- The sample is a subset of the (entire) population, in the case of the first example it is the persons we ask or interview in the second case it is the rainfall data we observe and collect. The sample represents a subset of still manageable size of the entire population. Samples are collected and statistics are calculated from the samples so that one can make inferences or extrapolations from the sample to the population. This process of collecting information from a sample is referred to as sampling.
- The best way to avoid an unrepresentative or biased sample is to select the observations (or individuals) randomly. A random sample is defined as a sample where the probability that any individual member from the population being selected as part of the sample is exactly the same as for any other individual member of the population.

Give it a thought / example for your own contemplation:

After a flood hitting the Tehsil/district you are asked to conduct a survey that will ultimately "describe the average disaster victim". Your boss already has prepared a list with the questions that he wants you to ask. You have to decide upon the strategy to obtain the answers. Three strategies are given as input, please select the most appropriate one to avoid unrepresentative sample and justify your decision:

a) You and your colleagues are positioning yourselves in front of the hospital, the morgue and the office of the Deputy Commissioner. You will ask every 10th person that is leaving those buildings about their personal suffering and loss.

b) You and your colleagues decide to interview only 150 persons. Since the town has 1,50,000 inhabitants you decide to interview every 1000th person. You get hold of an alphabetical list of all inhabitants and will put every possible effort in finding the selected persons.

c) Instead of obtaining the list of all registered inhabitants, which might be difficult since the officer–in–charge is totally absorbed by other emergency duties, a colleague of yours proposes to contact “Faircell” a local mobile phone operator and to draw the sample from their subscribers. Faircell is very enthusiastic about supporting the relief operations and proposes to you to limit the population wherefrom the sample is drawn to only those people that have used their number/mobile at least once over the last 5 days within the Tehsil. Thereby eliminating all those who have been absent while the disaster occurred.

Answer: b. is the most appropriate method. The sample is small (1 per mil) but satisfies the demand for arbitrary selection: each subject of the population faces the same probability to be chosen into the sample.

We will now focus on the description of rainfall using very basic statistical terms, some of which you might have heard about.

Nevertheless, we will see that this specific view, using simple tools of descriptive statistics to analyse rainfall data will already increase/anchor knowledge about disaster risks/natural hazards to a great extent and this could help us already in order to improve Disaster Risk Management.

4.3 Measures of central tendency and spread

In mathematics and statistics, an average, Central Tendency of a data set is a measure of the "middle" or "expected" value of the data set. The central tendency is essential in order to know what to expect – in our case, the average amount of rainfall in a particular time of the year.

The spread is important because rainfall as a statistical unit is a random variable and will be situated within a range, i.e. in one year it rains only 50mm in a particular month, in the next year it rains more the 400mm in the same month.

An underlying assumption (that can be proven) of the distribution of randomly captured data observed in real life is that the sum of a large number of independent random variables is distributed approximately normally.

There are many different instruments of descriptive statistics that can be chosen to measure the central tendency of the data items. These include arithmetic mean, the median and the mode. Other statistical measures such as the standard error / standard deviation, skewedness, kurtosis and the range are called measures of spread and describe how spread out the data is.

The distribution function shown in this page is the typical bell curve with the mean, median (and mode) at the peak and the standard deviation as the dark blue shaded area. A flatter curve would mean a higher spread (or larger standard deviation) and steeper curve would mean a lower spread (or smaller standard deviation). The bell shape above indicates that about 2/3 (68,2%) of the values fall within a range of one standard deviation below and above the mean, about 95.4% of the values are located within two standard deviation below and above the mean.

This of course holds true for a perfectly normally distributed population (which is quite likely) and what is less likely for a perfectly normally distributed sample,

Unfortunately there are many possible pitfalls when it comes to sampling but we will skip these (although come back later in module C) which might spoil our conclusion but we will postpone this discussion until later.

4.4 Average or mean and variable scales

The sample mean is an estimator available for estimating the population mean. This statement seems to be tautological, but if you read carefully about sample and population you notice that is not. It is a measure of location, commonly called the (arithmetic) average.

Its value depends equally on all of the data which may include outliers. It may also be biased; the sample of rainfall data might differ substantially from the population. It may not appear representative of the district or region for skewed data sets.

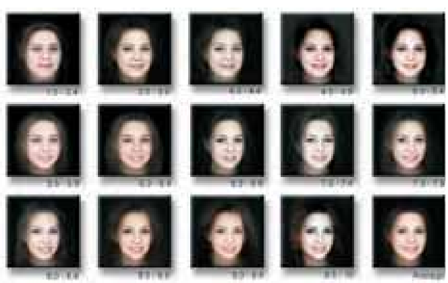


Figure B4.1: These women do not exist. Each of them is a composite of about 30 faces that were created to find out an average of good looks on the Internet. Copyright: Flickr/ manitou2121

The sample mean is however the most important indicator of the expected mean of the population. In future we will use the word mean synonymously with average.

The sample (arithmetic) mean is calculated by taking the sum of all the data values and dividing by the total number of data values.

In statistics one differs between different types of variables of the following scales:

Nominal scale

At the nominal scale, one uses labels; for example: colours. Nominal measures offer names or labels for certain characteristics. Variables assessed on a nominal scale are called categorical variables; the central tendency of a nominal attribute is given by its mode; neither the mean nor the median can be defined.

Give it a thought / example for your own contemplation:

Over the last 10years a particular district has been hit by 5 earthquakes, 2 drought periods and 3 floods. In total, 10 major disasters occurred in this particular district.

a) The average disaster is none, because you cannot calculate an average of distinct nominal values.

b) On average, over the last 10 years, each year one disaster occurred.

c) The most frequent type of disasters in this district is earthquakes.

Ordinal scale

In this scale type, the numbers assigned to objects or events represent the rank order (1st, 2nd, 3rd etc.) of the entities assessed. An example of ordinal measurement is the results of a race. The central tendency of an ordinal attribute can be represented by its mode or its median, but the mean cannot be defined.

Give it a thought / example for your own contemplation:

a) A survey conducted among young motorists revealed that they are more afraid of crime than of natural disasters. Accidents ranked least on the list of things young motorists are afraid of.

b) The five biggest disasters in India over the last ten years, according to people affected:

2002	Drought in UP and other states	300 million
2000/01	Drought in Gujarat and other states	50 million
2002	Flood in Assam and other states	42 million
2004	Flood in Dabhanga and other areas	33 million
2000	Flood in Birbhum and other areas	25 million

c) The ranking used in "b" may not a useful. The numbers are too big for an accurate count. The definition for somebody to be affected may vary. Is it better to use this ranking: no there are two different purposes of these statistics.

2001	Earthquake Gujarat	20,005
2004	Tsunami in Tamil Nadu, etc.	16,389
2005	Earthquake in Kashmir	1,309
2003	Heat wave in Andhra Pradesh	1,210
2005	Flood in Gujarat and other states	1,200

Interval scale

Quantitative attributes are all measurable on interval scales, as any difference between the levels of an attribute can be multiplied by any real number to exceed or equal another difference. A familiar example of interval scale measurement is temperature but also rainfall quantity. **Ratios** between numbers on the scale are not meaningful, so operations such as multiplication and division cannot be carried out directly. But ratios of differences can be expressed; for example, one difference can be twice another. The central tendency of a variable measured at the interval level can be represented by its **mode**, its **median**, or its **arithmetic mean**. Statistical dispersion can be measured in most of the usual ways such as **range**, **inter–quartile range**, and **standard deviation**.

Further examples:

Beaufort scale[1] for measuring wind

One measure of malnutrition: For reporting nutrition levels are aggregated into classes of malnutrition. The scale used is the Body Mass Index (BMI[2]) Mild Malnutrition (BMI=17–18.5) Moderate Malnutrition (BMI=16–17) Severe Malnutrition (BMI below 16).

4.5 Median, quartiles and mode

A median is described as the numeric value separating the higher half of a sample, a population, or a probability distribution, from the lower half. If the two most central values are distinct from the median, the median is calculated adding these two and dividing by 2. The mode is the value that occurs the most frequently in a data set but it is not so important in the observation of continuous data like rainfall. More important are the **quartiles**. These are three values which divide the sorted data set into four equal parts, so that each part represents one fourth of the sampled population.

Logically the **median** is the second quartile. The quartiles are values to describe the spread of a distribution in simple terms. **Mode is** the most frequent number in a data set.

Comparison of common averages			
Type	**Description**	**Example for sample**	**Result**
Arithmetic mean	Total sum divided by quantity of integers	(1+2+2+3+4+7+9) / 7	4
Median	Middle value that separates the greater and lesser halves of a data set	1, 2, 2, 3, 4, 7, 9	3
Mode	Most frequent number in a data set	1, 2, 2, 3, 4, 7, 9	2

[1] http://en.wikipedia.org/wiki/Beaufort_scale (last access 23.8.2013)

[2] The Body Mass Index (BMI) is defined as weight in kg divided by the square of the height in meter. Source for definition of malnutrition cut–off points: World Food Programme (2002): Emergency Field Operations Pocketbook

4.6 Measures of spread: range and standard deviation

In descriptive statistics, the range is the length of the smallest interval which contains all the data. It is calculated by subtracting the smallest observation (sample minimum) from the greatest (sample maximum) that is the difference of maximum – minimum and provides an indication of statistical dispersion. It is measured in the same units as the data. It is the most simple and so least indicative measure of dispersion.

Standard deviation is a widely used measure of the **variability** or **dispersion**; it shows how much **variation** there is from the "average" (mean). It may be thought of as the **average difference** of the scores from the **mean of distribution**, how far they are away from the mean. A low **standard deviation** indicates that the data points tend to be very close to the mean, whereas high **standard deviation** indicates that the data are spread out over a large range of values. Since it is also measured in the same unit as the data, it is usually preferred to the variance.

The **standard deviation** of a statistical population or a data set is the square root of its **variance**. The **variance** is the square of the **deviation** of that statistical population from its **mean** (see formula and / or sample calculation above).

The term **standard error** is derived from the fact that the standard deviation of the error (the difference between the estimate and the true value) is the same as the standard deviation of the estimates themselves; this is true since the standard deviation of the difference between the random variable and its expected value is equal to the standard deviation of a random variable itself.

In many practical applications, the true value of the standard deviation is unknown. As a result, the term standard error is often used to refer to an estimate of this unknown quantity.

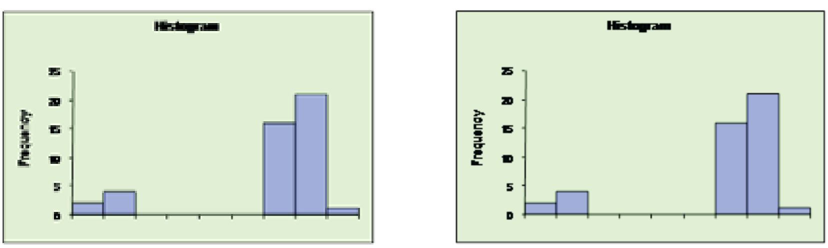
For the purpose of simplicity we will stick to the term standard deviation, knowing it is an estimate.

Comparison of measures of spread			
Type	**Description**	**Examples for sample**	**Result**
Range	Highest value – Lowest Value	1, 2, 3, 4, 4, 5, 6	5
Variance	Square of the deviation of that variable from its expected value or mean	1, 2, 2, 3, 4, 7, 9 the mean is 4 so the variance $\partial^2 = \frac{1}{(N-1)}\sum_{i=1}^{N}(X_i - \bar{X})^2$ or $\frac{1}{6}\sum_{i=1}^{7}(X_i - 4)^2 = 2{,}66$	2.66
Standard Deviation	The square root of Variance	1, 2, 2, 3, 4, 7, 9	1.63
Kurtosis	Peakedness	1, 2, 2, 3, 4, 7, 9	–0.64
Skewedness	Tailedness	1, 2, 2, 3, 4, 7, 9	–0.17

4.7 Other measures of spread

Skewedness is a measure of the asymmetry of the probability distribution.

- Positive skew: The right tail is longer; the mass of the distribution is concentrated on the left of the figure. It has relatively few high values. The distribution is said to be right–skewed. Example – see below.
- Negative skew: The left tail is longer; the mass of the distribution is concentrated on the right of the figure.



Negative skew **Positive skew**

Figure B4.2: Skewedness

It has relatively few low values. The distribution is said to be left–skewed. Example (Figure B4.2)

What does this mean as we look at the rainfall?

- With a positively skewed distribution we can reasonably expect higher rainfalls although the expected mean is low.
- With negative skew we can reasonably expect lower rainfalls although the expected mean is high.

Kurtosis is a measure of the "peakedness" of the probability distribution of a real–valued random variable. Higher **kurtosis** means more of the variance is the result of infrequent extreme deviations (it is "peaked") as opposed to frequent modestly sized deviations (it is more spread out) (Figure B4.3).

What does this mean as we look at the rainfall?

- With a rainfall distribution with a positive kurtosis we can expect with higher probability that the expected rainfall will be located close to the expected mean.
- With negative kurtosis we can reasonably expect wider spread of the rainfall data.

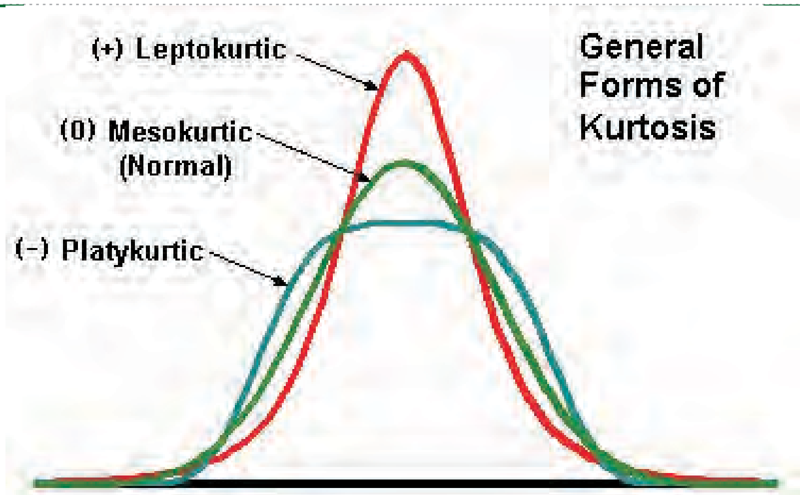


Figure B4.3: Kurtosis

Range is the difference between the highest value and the lowest.

4.8 Rainfall figures from Bihar State

Let us look at the rainfall data of Bihar using these simple statistical indicators employing the built–in formulas of EXCEL (version 2003 used here for demonstration).

Let's look at the **EXCEL Template** and the sheet: PATNA – template.

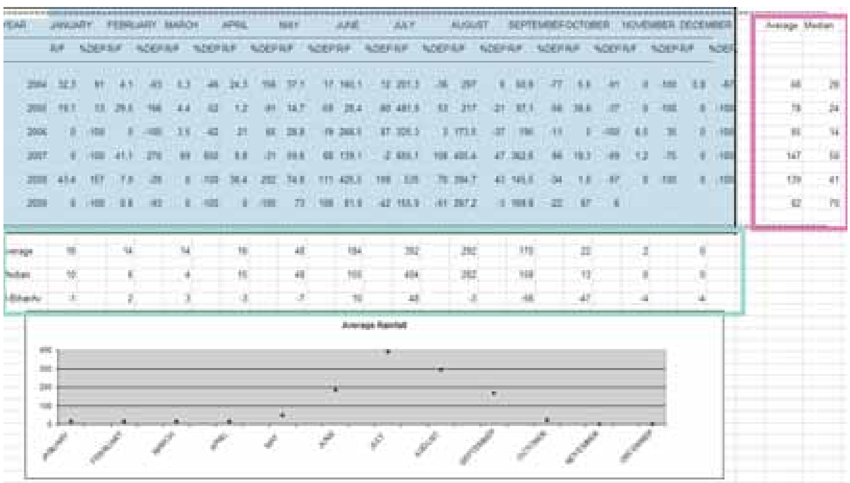


Figure B4.4: Average monthly rainfall

Source: Own compilation

Rainfall figures from Bihar State – example of mean and median

You will remember the file from unit 3. Now look at the two columns at the outer right, the pink rectangle:

You will see the average **mean** and the **median** for annual rainfall figures: In 2007 we had the highest **average** and this is usually in our case the better figure than the **median** because the mean tells you how much rain (has fallen on average per month. Note that the years with most rainfall were 2007 and 2008, i.e. 147 and 139 respectively. You might also notice that the year with the highest average rainfall does not correspond with the year when median rainfall reached its maximum (2007 vs. 2009).

Average and **median** are calculated in the template with the EXCEL formulas:

AVERAGE (<range>) and MEDIAN (<range>), the graph on the bottom shows the rainfall pattern of monthly averages over the last 6 years.

Apart from this we have a distinct monsoon pattern of rainfall: highest precipitation starting in June and ranging until September, very low rainfalls for the rest of the year.

In the next module we will see what these figures in detail will be able to tell us about possible floods and rains in the past.

What can we learn of the measures of central tendency?

Mean or **average** on monthly basis show the monsoon pattern of rain. Rain starts in May and ends almost completely in October, at least in some districts. Looking at the rainfall files we used so far, it seems that rainfall /a specific area's rainfall may not be the **only** cause for floods in the area. An extensive network of rivers may contribute to downstream inundation due to upstream inflow.

When are the different measures of central tendency best used?

Mean gives a better understanding in the annual figures. The **median** here is usually significantly lower due to the fact, that only three months of the year have abundant rainfall. However measure of spread will be looked at in the next module and will allow another fresh look at the significance of the mean.

How are the different measures of central tendency applied for DRM?

Mean and **median** in our example give us evidence of rainfall pattern throughout the year and when to expect heavy rainfall, not much more but also not much less. Since abundant rainfall is the main reason for floods and the lack of rain is the main reason for drought, the two disaster types with –usually– the largest number of affected, rainfall data is a prerequisite for a thorough statistical analysis in order to monitor the situation and to take appropriate action when is too low or too high.

4.9 Guided exercise (1)

1. Look at the Map of Bihar and identify districts you want to know more about their rainfall patterns (http://imd.gov.in/section/hydro/distrainfall/districtrain.html).

2. Select the district you want to find out about its rainfall –> MUZAFFARPUR, because it is a district in the Gandak River Area, likely to be affected by heavy rainfall (http://imd.gov.in/section/hydro/distrainfall/webrain/bihar/MUZAFFARPUR.txt).

3. Go to the EXCEL template, Copy the sheet PATNA–template in the same file (Edit/Copy Sheet, select <Copy sheet>).

4. Rename sheet created PATNA–template (2) to MUZAFFARPUR (Click on Tab and rename). Replace in Cell A6 PATNA with the MUZAFFARPUR data.

5. Select from 2.) Data for years 2004 – 2009 (see picture – Click to enlarge) from IMD Website; select only the rainfall data as indicated in adjacent picture (Screenshot 1 – Figure B4.5 – Enlarge if necessary).

6. Delete old content before (Delete A19:Y29).

7. Copy to EXCEL template (Paste to Cell A19 = '2004').

8. Expand text (Copied content from WEBPAGE was text) (select cells as indices, convert to cells: <Data/Text to Columns>).

Figure B4.5: Screenshot 1

Source: Own compilation

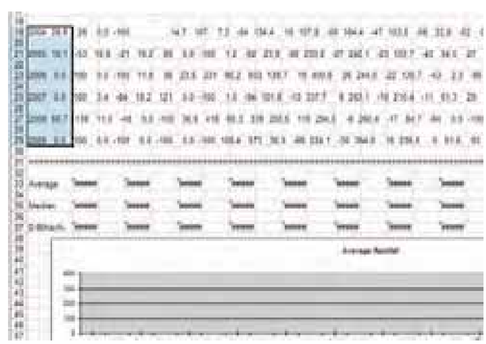


Figure B4.6: Screenshot 2

Source: Own compilation

9. Control and eventually adjust Graph (Place cursor on Graph: Graph/Data Source/Modify in Values of Series “Average Rainfall” MUZAFFARPUR for PATNA–Template). This might not be necessary, if copying has been done correctly.

10. Well done and finished.

5

LEARNING UNIT B–5: MORE ON MEASURES OF DISPERSION

5.1 What is this unit about?

In this unit we will explore the statistical terms of central tendency and of dispersion in practice. While the "Guided Exercise" is still ongoing we are calculating the statistics (that is to say we are applying the copied formulas of our master excel file) of another district of our choice.

In this unit you will also be introduced to StatistiXL. If you do not wish to use this software, or if you are using OpenOffice software instead of Microsoft Excel, we have indicated the corresponding alternatives for you at the end of this unit. For details about StatistiXL, please refer to the text box below:

Expected results

The participants will apply the statistical terms and grasp their significance in application. The second part of the "Guided Exercises" teaches how to apply the add–ins or statistical analysis functions for another selected district. It is assumed that participants understand how to apply this procedure for any district chosen in India (where data are available), how we chose and why we chose this scenario in Bihar and why they use fact based information for the benefit of Disaster Risk Management. The EXCEL Add–ins and / or statistical analysis functions are introduced and participants can understand its content and draw.

5.2 Examples of measures of dispersion – rainfall figures from the Bihar State

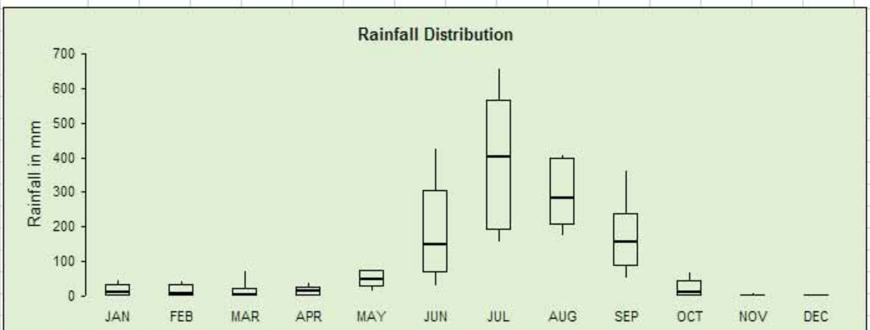
Let us look at the rainfall data of Bihar again. Now we will use the appropriate statistical indicators. First of all, you will have to download a full working version of StatistiXL. You can download it here. StatistiXL includes all Statistical modules, Help files and Example files that you will need. It will run on your system as a test version for 30 days and allows you to check out all of its features by applying your own data. After the 30 days have elapsed, you will have to purchase a license from the online store if you want to continue using StatistiXL (http://www.statistixl.com (last access 23.8.2013).

Once you have installed StatistiXL, let us look at the **EXCEL Template2** and the sheet: PATNA – template. Click the StatistiXL Menu right of the help menu:

	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
2004	32	4	5	24	37	160	201	297	51	6	0	1
2005	19	30	4	1	15	28	482	217	97	40	0	0
2006	0	0	4	21	29	267	325	174	195	0	7	0
2007	0	41	69	9	60	139	655	405	363	19	1	0
2008	43	8	0	38	75	426	535	395	146	2	0	0
2009	0	1	0	0	73	82	156	267	170	67	0	0
Average	16	14	14	16	48	184	392	292	170	22	2	0

Descriptive Statistics for:
Input Range = PATNA-Template!$B$53:$M$59

	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
Mean	15,8	13,9	13,7	15,6	48,0	183,6	392,4	292,5	170,1	22,3	1,3	0,1
Median	9,6	6,0	4,0	14,9	48,4	149,6	403,6	282,1	157,7	12,6	0,0	0,0
Std Error	7,7	7,0	11,1	6,1	10,1	58,4	80,5	38,2	43,9	10,8	1,1	0,1
Std Dev.	18,9	17,2	27,2	15,0	24,8	143,1	197,1	93,5	107,6	26,4	2,6	0,3
25th Perce	0,0	0,6	0,0	0,9	25,3	68,5	190,0	206,1	85,5	1,4	0,0	0,0
75th Perce	35,1	32,4	21,2	27,8	73,5	306,3	565,0	397,4	236,9	46,5	2,5	0,2
Minimum	0,0	0,0	0,0	0,0	14,7	28,4	155,9	173,5	50,8	0,0	0,0	0,0
Maximum	43,4	41,1	69,0	38,4	74,8	425,5	655,1	405,4	362,6	67,0	6,5	0,8
Skewness	0,6	1,1	2,4	0,5	-0,2	1,0	0,0	0,1	1,2	1,1	2,3	2,4
Kurtosis	-1,7	-0,8	5,9	-1,0	-2,0	0,7	-1,7	-1,6	2,1	0,3	5,3	6,0



ARARIA / AURANGABAD / GAYA / PATNA / **PATNA-Template** / LT_Average

Figure B5.1: Rainfall figures from the Patna District, Bihar State

Source: Own compilation

5.3 Guided exercise (2)

1. Go to the EXCEL template2 and open the sheet PATNA–template.

2. Make sure the Section (A53:M61) from the sheet “PATNA–template” has been copied to the sheet “MUZAFFARPUR”, make sure you copy it to the correct cell in this sheet, that is A53. The copied data will then contain the rainfall data of the current sheet, but in reduced form (without differences from averages %DEP). This is the way we will need them for the statistical analysis.

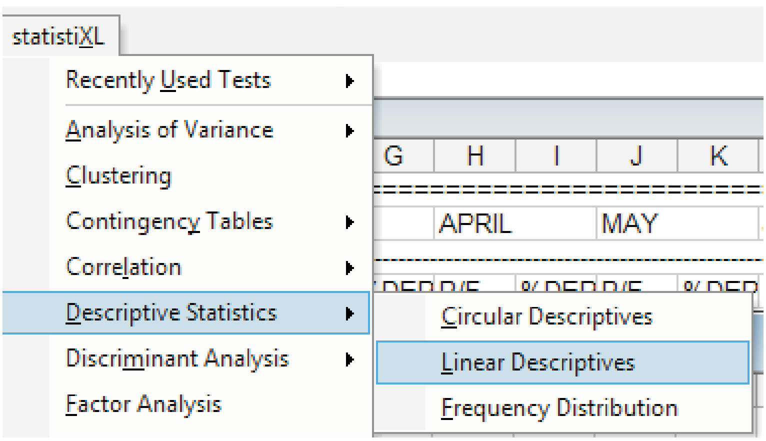


Figure B5.2: Screenshot of the StatistiXL on descriptive statistics.

Source: Own compilation

3. Go to Select the menu StatistiXL, then (Descriptive Statistics/ Linear Descriptive).

4. Then select the non–empty variable range (B53–M58) (the rainfall data, labels – the name of the months– in first row of newly copied table – row53 of sheet “MUZAFFARPUR”) and the output range, where the data will be displayed (e.g. A63) as in Figure B5.3a. Don’t press <Enter>.

5. Select the indicated statistics from the STATISTCS Tab. Don’t press <Enter>.

6. Select the Box and Whiskers Plot from the PLOT Tab, then finish by pressing OK or press <Enter>.

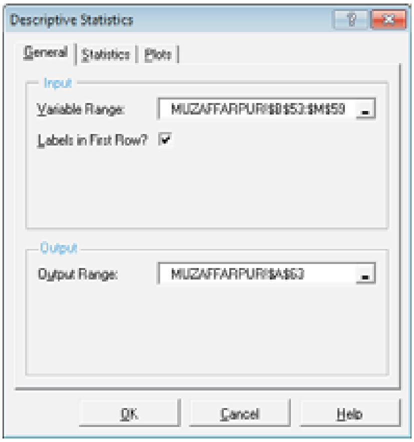


Figure B5.3a

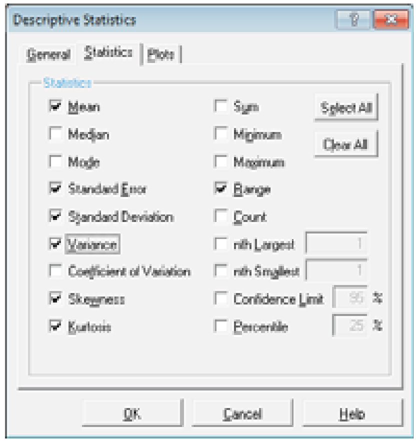


Figure B5.3b

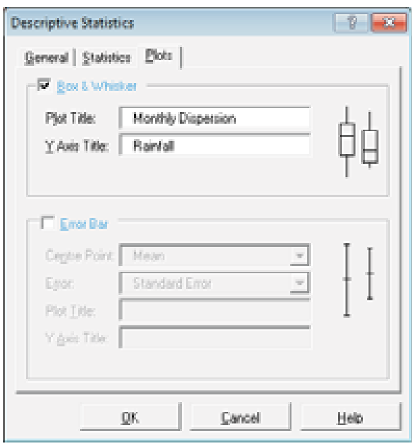


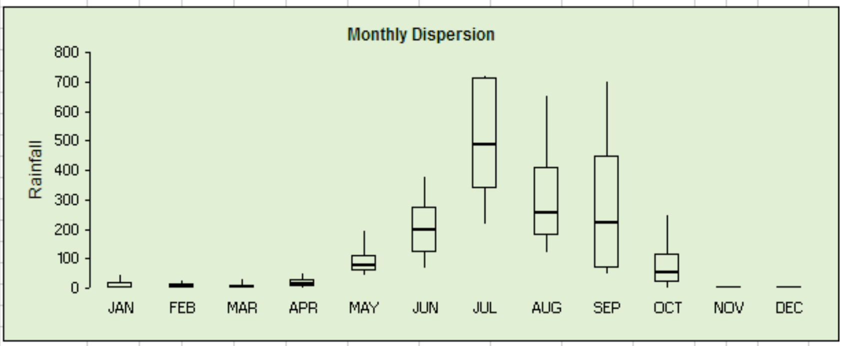
Figure B5.3c

7. You can do a bit of formatting of the cells of the resulting table (B66:M78) and you can move the graph like this.

	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NO'	DEC
2004	44	0	0	50	79	229	576	199	46	2	0	0
2005	14	12	3	15	42	141	400	251	77	53	0	0
2006	0	0	2	22	72	242	378	122	365	26	5	0
2007	0	24	29	9	81	167	719	649	700	247	1	0
2008	0	4	0	13	61	377	712	332	277	54	0	0
2009	0	2	2	0	195	68	216	262	165	71	0	0
Average	10	7	7	18	88	204	500	303	272	75	1	0

Descriptive Statistics for:
Input Range = MUZAFFARPUR!$B$53:$M$59

	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
Mean	9,5	7,0	5,9	18,2	88,2	203,8	500,2	302,6	271,8	75,3	0,9	0,0
Median	0,0	2,8	1,7	14,2	75,5	197,6	488,0	256,7	221,3	53,3	0,0	0,0
Std Error	7,2	3,8	4,6	7,1	22,1	43,2	82,5	75,0	98,7	35,7	0,7	0,0
Std Dev.	17,5	9,3	11,4	17,4	54,1	105,8	202,2	183,7	241,9	87,5	1,8	0,0
Variance	306,9	86,8	129,2	301,2	2927,6	11192,5	40882,4	33733,1	58503,7	7654,0	3,2	0,0
25th Perce	0,0	0,2	0,0	6,5	56,4	122,7	337,3	179,6	69,0	19,8	0,0	0,0
75th Perce	21,2	15,1	9,7	28,8	109,1	275,5	714,0	411,1	449,0	115,0	1,9	0,0
Minimum	0,0	0,0	0,0	0,0	42,0	67,8	216,1	122,2	46,0	1,6	0,0	0,0
Maximum	43,5	23,6	29,0	50,4	194,8	377,3	719,4	649,3	700,0	246,8	4,5	0,0
Range	43,5	23,6	29,0	50,4	152,8	309,5	503,3	527,1	654,0	245,2	4,5	0,0
Skewness	2,0	1,5	2,4	1,5	2,1	0,6	-0,2	1,7	1,3	2,0	2,2	-
Kurtosis	3,9	1,4	5,8	3,0	4,7	0,8	-1,6	3,4	1,5	4,5	4,9	-



MUZAFFARPUR / ARARIA / AURANGABAD / GAYA / PATNA / PATNA-Template / LT_Average

Figure B5.4: Rainfall figures from the Bihar State

Source: Own compilation

8. Well done and finished.

If you do not have StatistiXL available on your system, you may try some functions of Excel. In EXCEL go to: Functions: (Extra/Add–ins/Analysis (ticked)) –> Then use Extra/Analysis/<Stats function of your choice>.

For the above mentioned stats functions you can use the standard EXCEL functions (except range which is MAX –MIN), however there is no Boxplot, which would be nice to have for first exploration of data.

In case you are not using Microsoft Excel but OpenOffice instead, we have prepared the following table for you delineating the differences in command syntax for MS Excel and OpenOffice:

Table B5–1: Open office syntax for statistical analysis

MS Excel Function	Description	OpenOffice Syntax
AVERAGE	Returns the average of its arguments	AVERAGE(number_1; number_2; ... number_30)
KURT	Returns the **kurtosis** of a data set	KURT(number_1; number_2; ... number_30)
MAX	Returns the **maximum value** in a list of arguments	MAX(number_1; number_2; ... number_30)
MEDIAN	Returns the **median** of the given numbers	MEDIAN(number_1; number_2; ... number_30)
MIN	Returns the **minimum** value in a list of arguments	MIN(number_1; number_2; ... number_30)
MODE	Returns the **most common value** in a data set	MODE(number_1; number_2; ... number_30)
QUARTILE	Returns the **quartile** of a data set	QUARTILE(data; type)
SKEW	Returns the **skewedness** of a distribution	SKEW(number_1; number_2; ... number_30)
STDEV	Estimates **standard deviation** based on a sample	STDEV(number_1; number_2; ... number_30)
VAR	Estimates **variance** based on a sample	VAR(number_1; number_2; ... number_30)

Before discussing the results, let us look at the different measures of dispersion and their significance in this context.

5.4 How to use measures of dispersion for the analysis of your data?

We have used the measures **quartiles** (25%, Mean, 75%) as statistics. The values indicated tell us that 25% of the rainfall values are below the 25%, half are below the **median** and 75% below the 75%. These values are found in the table. The values are graphically displayed in the Box and Whiskers Plot and make the dispersion of the data easier to grasp. The box represents the 25%, 75% limits, the fat line in the centre is the **median.** Apart from this we see the **kurtosis** and the **skewedness** as well as the **standard deviation** (and its offspring: Variance etc.) for each month. The **standard deviation** is graphically shown by the box length of the box plot. Similarly the kurtosis is represented by the whole spread of the box + whiskers. A positive high **kurtosis** is graphically shown as a short graph as opposed to a long graph for a negative or low **kurtosis**. A positive **skewedness** is represented by a positive whisker (above the box), a negative **skewedness** is represented by a long whisker below the box, If both whiskers are of equal length than this would result in a low **skewedness**, neither positive nor negative.

When are the different measures of dispersion best used?

We can use the distribution measure to recognize the pattern especially of monsoon season rainfalls, since in the rest of the year there is almost no rainfall at all. Looking at the figures or the graph to tell where we have dispersions differing among the monsoon months. Negative **skewedness** will let us expect likely deviations to lower rainfalls, a high **kurtosis** tells us about relatively stable rainfalls in this month and so forth.

Where are the different measures of dispersion applied for DRM?

Disaster Risk Management cannot purely rely on the data about expected rainfall and the simple averages that we calculated in the previous unit. For a more profound analysis of rainfall date the so called measures of dispersion are essential because they allow you to conclude whether new observations are still in a range of “expected values” or if it is indeed a “strange” figure/value/observation that needs to be watched closely in order to be prepared. For instance, if we observe a high deviation to the positive side in our rainfall data, this would mean that even more excessive rain than having already experienced over the past years is likely to occur one day and that adequate preparedness and mitigation measures should be taken in order to respond appropriately to this scenario.

What else could we have learned from the rainfall figures from Bihar State?

Bihar is facing a serious drought arising out of scanty rainfall during the 2009 South West Monsoon whose arrival was 2 weeks late. The rainfall deficit in June was 62%. The State Government has declared 26 district as drought affected and started relief measures. The situation in the remaining 12 districts is also being monitored continuously and some of these districts may also be declared drought hit if the situation in these districts worsens. The names of 26 districts which have been declared drought affected, are in the following map of Bihar, (Figure B5.5):

Figure B5.5: Drought prone districts in Bihar

Source: Drought 2009, overview and management, www.agricoop.nic.in

5.5 Resume of the examples of rainfall figures from Bihar State using statistics

Let us sub sum the results for MUZAFFARPUR from the previous unit and add the findings using measure of dispersion. First we will be able to relate the rainfall figures to flood only, risk of drought will be faced as a result of delay in rainfall during June 2009. The state received only 736.3 mm rainfall against 1039.2 mm with a deficiency of 29%.

Results and preliminary analysis based on measures of central tendency: Heavy rains in 2007, abnormal rainfall highs in July 2007 and 2008 (in next module we will look at historic disasters and possible reasons for them). MUZAFFARPUR has 6 months of almost no rain. Vulnerable months are probably July and August. But look at the extremely low rain in June 2009 followed by rainfall in July measuring only half of the average. What would be the consequences?

5.5.1 Results and preliminary analysis based on measures of dispersion

The months of May to October show considerable differences in the rainfall patterns. It is clear the expected rainfall rises to a peak in July, but August and September in MUZAFFARPUR can also have rainfalls that can match the precipitation in July. Preparedness for floods in these months should certainly not weaken due to the expectation that the rainfall peak has been reached in July.

We see further more reliable patterns of rainfall prediction in June and August, indicated by a rather low standard deviation compared to higher ones in July and August. This is immediately illustrated by the larger boxes in the Box & Whiskers graph for these months.

The dry months will be left out of consideration because rainfall figures are too insignificant to be compared with good reasoning.

You should be able to repeat this exercise for another district (like e.g. ROHTAS) to improve your expertise in managing the EXCEL analysis features and the interpretation of results for further districts.

Now it is up to you to practice and make use of what you have learned so far in order to become more familiar with the use of statistics!!

5.5.2 Independent exercise of rainfall figures from Bihar State using statistics

1. Use a district of your choice in the state of Bihar.
2. Use the Templates and calculate Indicators of Central Tendency and Dispersion.
3. Observe the descriptive statistics, write down your findings.
4. Put this result (no EXCEL file needed) in few words on the discussion forum under: “Solutions of independent exercise Module A”. You may solve this exercise in groups or on your own. If you solve it in a group, make sure the name of the group members will appear.


ases For Databases and tics for Managemtr
Statistics Management Disaster Dadabas

6

SELF – ASSESSMENT MODULE B

Question 1

What can data tell us?

Which are questions that can be answered with the help of statistics?

Tick selection (or comment). Sometimes more than one selection is correct, sometimes none.

Select the correct answer

a) What were the climatic conditions in the past?

b) When will a disaster (like flood) strike my village?

c) The monsoon rainfall in one district generates usually higher rainfall than in another.

d) The next year will bring same monsoon rainfall as the average rainfall in the past five years in this district.

Question 2

Measures of central tendency

Which of the following indicators are statistical measures of central tendency?

Tick selection (or comment). Sometimes more than one selection is correct, sometimes none.

Select the correct answer.

a) Mean

b) Prejudice

c) Bias

d) Average

e) Range

f) Trend

Question 3

Which of the following indicators are statistical measures of dispersion?

Tick selection (or comment). Sometimes more than one selection is correct, sometimes none.

Select the correct answer.

a) Standard deviation

b) Count

c) Variance

d) Minimum

e) Skewedness

Question 4

The population and the sample

Which of the following statements are true?

Tick selection (or comment). Sometimes more than one selection is correct, sometimes none.

Select the correct answer.

a. A sample is a subset (part) of the population

b. Since the Indian population is so big, a population sample to draw conclusions from has to be very big.

c. Reliable, scientific and random sampling is the precondition for any statistical inference.

d. The most important criteria of a reliable sample is the size.

Question 5

Which is a random sample?

Of the following examples, please indicate the random sample?

Tick selection (or comment). Sometimes more than one selection is correct, sometimes none.

Select the correct answer.

a. Of all villages in a district, select those below 5000 inhabitants.

b. Of all villages in a district, select for different size groups (big, medium, small), a percentage difference to each, but for each group the villages are selected randomly.

c. Select a sample of all Indian citizens, picking only those taller than 5 feet.

d. Select a sample of all Indian citizens, picking from the Income tax Office

e. Of all disaster prone districts, choose two groups, the disaster prone district and the rest. Select an exhaustive record by counting (e.g. every thousandth) two samples.

Question 6

About Variable scales

There are variable scales, which determines and limit the use of variables. Which of the following statements are correct?

Tick selection (or comment). Sometimes more than one selection is correct, sometimes none.

Select the correct answers.

a. You cannot calculate a mean of values from ordinal scaled variable.

b. Variable and values means the same.

c. A variable `sex` may have the value `male`.

d. Age is an interval scaled variable.

e. An interval scaled variable is always `ordinal`, allowing values of this variable to be ranked.

f. To calculate an arithmetic mean, the variable does not have to be of interval scale.

Question 7

Measure of Dispersion

Which of the following statements are correct?

Tick selection (or comment). Sometimes more than one selection is correct, sometimes none.

Select the correct answer

a. Negative skew means that the distribution of the values has a right tail

b. Kurtosis is a measure of `flatness` of the curve of the distribution of the values.

c. A higher positive kurtosis of rainfall means that the rainfalls can be expected with the higher probability to be in the area around the peak than if the kurtosis is smaller or negative.

d. A boxplot can indicate skewedness as well as kurtosis if not as well as a curve of a distribution.

Question 8:

Interpretation of statistical measure for DRM

Which statements derived from statistical analysis of district rainfalls are correct?

Tick selection (or comment). Sometimes more than one selection is correct, sometimes none.

Select the correct answers.

a. You can predict rainfall for the next monsoon base on the rainfall figures of previous years.

b. You can compare rainfall patterns in different districts and assume from these patterns a higher risk of floods in one compared to another.

c. Kurtosis and skewedness tell us something about the pattern of monsoon rainfall like post monsoon rainfalls have been frequent in ascertain district in the past

d. Rainfall in the past has nothing to do with rainfall in the future

e. The risk of flood is not only due to rainfall, many other factors influence the occurrence of floods: drainage, irrigation schemes, upstream events, etc.

Answers

1. a & c
2. a &c
3. a, c & e
4. a & c
5. b & d
6. a,c,d & e
7. c & d
8. b, c & e

7

BIBLIOGRAPHY MODULE B

BMTPC (2006). Vulnerability Atlas of India; Building Material Technology Promotion Council, New Delhi

ISDR. International Strategy for Disaster Reduction (http://www.unisdr.org/) (last access 23.8.2013).

Scott Ch. (2005). Measuring Up to the Measurement Problem. The role of statistics in evidence based policy–making– Paris 21 London School of Economics.

MHA (2004). Disaster Management In India, Government of India Ministry of Home Affairs.

World Food Programme (2002): Emergency Field Operations Pocketbook.

EM–DAT. The International Disaster Database. Centre for the Research on the Epidemiology of Disasters (http://www.emdat.be/) (last access 23.8.2013).

SwissRE (2/2009). Natural catastrophes and man–made disasters in 2008. Sigma Report, (http://media.swissre.com/documents/sigma2_2009_en.pdf) (last access 23.8.2013)

SwissRE (1/2010). Natural catastrophes and man–made disasters in 2009. Sigma Report, (http://media.swissre.com/documents/sigma1_2010_en.pdf) (last access 23.8.2013)

Drought Management Division of Ministry of Agriculture and Cooperation (2009). Overview and Management Strategies.

Module C

Use Statistical Information on Recent and Past disasters

Contents

Module C

1	**Learning Unit C–1: India and the international scenarios**	115
	1.1 What is this module and this unit about?	115
	1.2 Scenario 1 – Cyclones	118
	1.3 Independent exercise: Cyclones	121
	1.4 Scenario 2 – Earthquakes	121
	1.5 Independent exercise: Earthquakes	123
	1.6 Scenario 3 – Floods	124
	1.7 Independent exercise: Floods	125
	1.8 Independent exercise	126
	1.9 Conclusion from exercises	127
2	**Learning Unit C–2: Frequency and impacts of major disasters in India**	129
	2.1 What is this unit about?	129
	2.2 Frequency and impacts of disasters in states of India: A comparison	130
	2.3 Conclusion from this unit	135
3	**Learning Unit C–3: Selected areas of disasters in India and their frequency and impacts**	137
	3.1 What is this unit about?	137
	3.2 Guided and independent exercise	138

3.3 History of natural disasters in Andhra Pradesh 139
3.4 Guided exercise on cyclones in Andhra Pradesh 140
3.5 Independent exercise: Disasters in Orissa and Tamil Nadu 143
3.6 Andhra Pradesh: Conclusion from exercise 144
3.7 Gujarat: History of disasters 145
3.7.1 Gujarat: More on earthquakes 147
3.8 West Bengal and Bihar: History of natural disasters 149
3.8.1 West Bengal and Bihar: Details on floods 150
3.8.2 Guided exercise: Floods in Bihar 151
3.8.3 Independent exercise: Floods in Bihar 151
3.8.4 West Bengal and Bihar: Conclusions on flood disasters 152
4 Self–Assessment for Module C 155
5 Bibliography Module C 161

1

LEARNING UNIT C–1: INDIA AND THE INTERNATIONAL SCENARIOS

Aim of the module is to prevent and mitigate impacts of disasters by increasing awareness of frequencies of disasters through better knowledge of patterns if they exist. Improved information can lead to a regional and local early information system and improved preparedness and also to more appropriate post–disaster reactions.

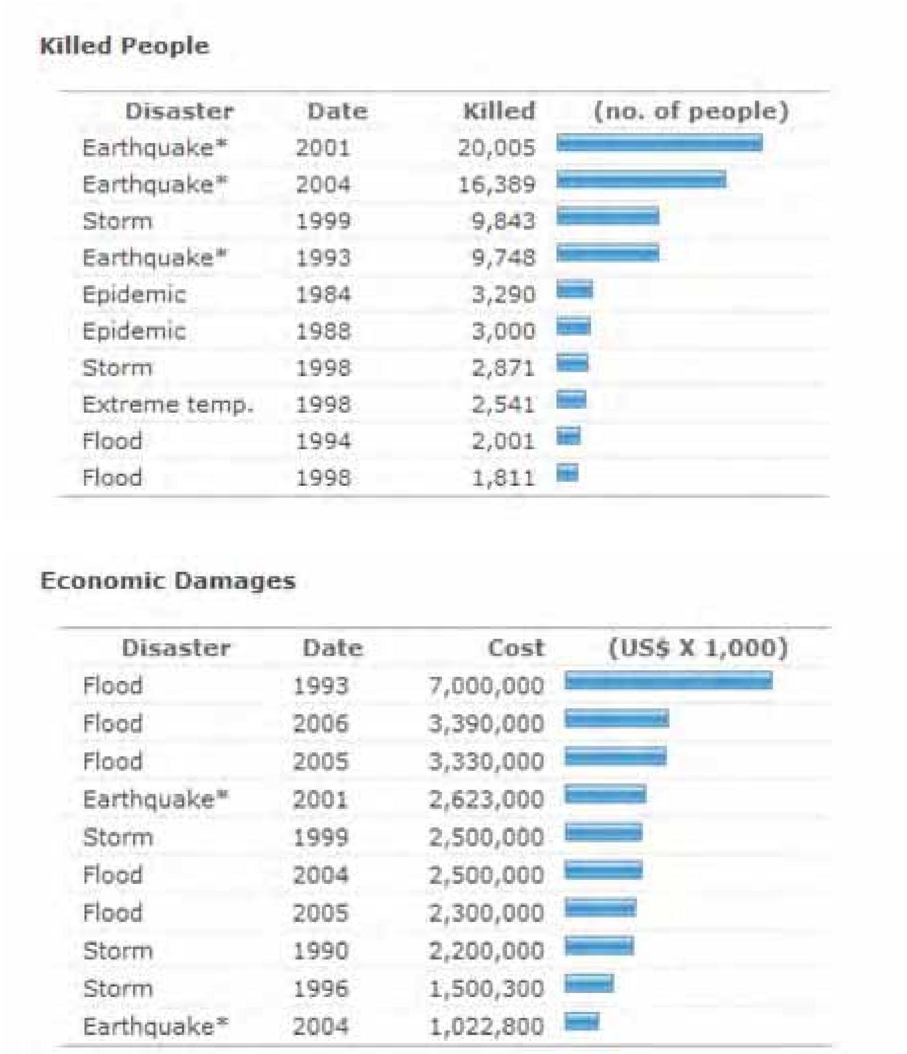
1.1 About the Module

In Module B, the risks for the very limited area of a few districts in Bihar were analysed based on Statistical data. The Module C embarks more on the national and international perspective with the purpose to point at challenges and best practices at national and international level.

The Module C will look at past disasters under statistical aspects. This means, we will analyse them under measurable (i.e. quantitative) aspects and answer the following questions:

- Which disasters had the highest risks, i.e. which ones affected people most and which caused the highest economic damages?

  For this purpose, we will again select typical types of natural disaster scenarios: Cyclones, earthquakes and floods. Look at the following stats on the number of people killed and the costs of the (economic) damages due to disasters in India between 1980 and 2008 (Figure C1.1):



Killed People

Disaster	Date	Killed	(no. of people)
Earthquake*	2001	20,005	
Earthquake*	2004	16,389	
Storm	1999	9,843	
Earthquake*	1993	9,748	
Epidemic	1984	3,290	
Epidemic	1988	3,000	
Storm	1998	2,871	
Extreme temp.	1998	2,541	
Flood	1994	2,001	
Flood	1998	1,811	

Economic Damages

Disaster	Date	Cost	(US$ X 1,000)
Flood	1993	7,000,000	
Flood	2006	3,390,000	
Flood	2005	3,330,000	
Earthquake*	2001	2,623,000	
Storm	1999	2,500,000	
Flood	2004	2,500,000	
Flood	2005	2,300,000	
Storm	1990	2,200,000	
Storm	1996	1,500,300	
Earthquake*	2004	1,022,800	

Figure C1.1: Number of people killed and the costs of the (economic) damages due to major disasters in India between 1980 and 2004

Source/Copyright: Prevention Web

We clearly see that even though earthquakes resulted in highest number of deaths, economic losses are highest for floods and storms (cyclones).

- Which regions in India are the most likely to be affected by natural disasters?

For this purpose we will identify the zones of the country affected most by the three major natural hazards. We will have to look at repetition patterns, frequencies and severity of disasters. In Module C we will also statistically analyse some of the effects of best practices to cope with disasters in the past. For this we will consult available statistical information on prevention, mitigation and other phases of disasters.

We will start with a comparison of disaster risks in India and at international level for various disasters. And for the benefit of critics of statistics we will start with an EXAMPLE OF WEAK STATISTICAL DISPLAY (Figures C1.2 and C1.3 and explaining text).

On the CD, provided to you in the training, we provide you with a positive example of international comparison of disasters see EMDAT_Comparison2010_DistasterResults.pdf. The purpose of Module C will also be to find, read and interpret the frequency data for disasters and to relate it to the human and economic loss caused by the disasters. Finally, we will point to some specific aspects concerning disaster risks, like man–made influences on disasters or external threats (i.e. glacial lakes in neighbouring countries). Also the ranking of man–made disasters in India in relation to natural disasters will be discussed.

Expected results

The participants will understand statistical tools to describe the impacts of disasters in India over the past years. The purpose is to learn from the statistical description of risks and the best practices already gained by others in order to apply them for your own purposes. We will NOT go much further than making you aware of disaster frequencies, national and international best practices.

1.2 Scenario 1: Cyclones

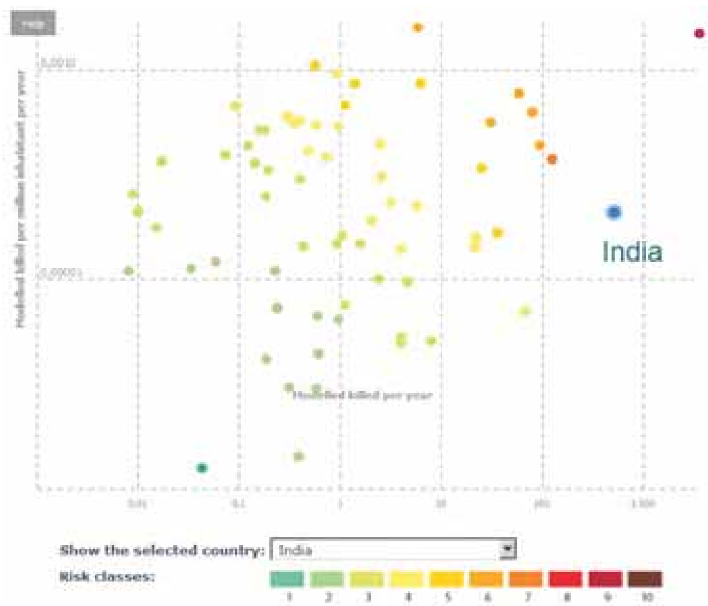


Figure C1.2: Mortality Risk Index (MRI) for tropical cyclones

Source: http://www.preventionweb.net/english/maps

Figure C1.2 shows a mortality risk index (that has been developed by UNEP's GRID initiative). On the x–axis it shows the absolute number of people killed according to the modelling assumptions of the authors. Obviously the numbers are in relation to the population of a country, otherwise they would not be comparable, which makes the graph a bit difficult to understand at first sight.

As you can see, India is the country with the second highest number of fatalities according to the modelling assumptions. On the y–axis the absolute number of fatalities is divided by the total population, i.e. it displays the risk of any person living in that country of getting killed due to a cyclone. As you can see, India, while it still has a high risk, it scores relatively well compared to other nations.

How come that India has the second but highest number of persons killed due to cyclones but compared to other countries, the risk of being killed is not that high? Well, it depends on the size of the population. It is important to note that the y–axis compares those killed by a cyclone with the entire Indian population i.e. it even takes into account all those who live in mountainous regions, without any risk for a cyclone.

To be more explicit: The only difference between the x– and the y–axis is that on the y–axis the numbers of the x–axis are divided by the population of the respective country. Now, look at the heading of the graph. It says: "Mortality risk index for tropical cyclones."

Reading this heading, would you expect that the graph depicts something about cyclones or about the population size? Consider that you live in a cyclone prone area. What factors are likely to determine the risk of getting killed? Maybe you would think about the location where you are when the cyclone strikes, the availability of shelter and maybe the strength of that shelter. Would you also consider the size of the population?

If you were to improve this graph, you might want to compare the number of persons killed with the absolute number of cyclones, the duration of cyclones, their intensity or the number of people actually exposed to the cyclone vs. the number of people injured or killed.

As you can see from this example, statistics does not so much depend on the actual calculation, or the method chosen for calculation, but on the primary assumptions of your model. It depends also on the variables you choose (persons killed, population size, number of cyclones, intensity, etc.) and second it depends on the relationship that you assume exists between the variables. For instance in Figure B3.3 the assumption is that for everybody living in the same country the risk of cyclones is the same. This might be a valid assumption with a very small country or a country with a similar (cyclone) risk in all parts of the country. But for a country like India it might be misleading.

We would encourage you how do you think the graph would look like if the unit of analysis were not countries but the Indian States and Territories, or if it were even the districts within a state if you are using similar analysis for India at state and district levels.

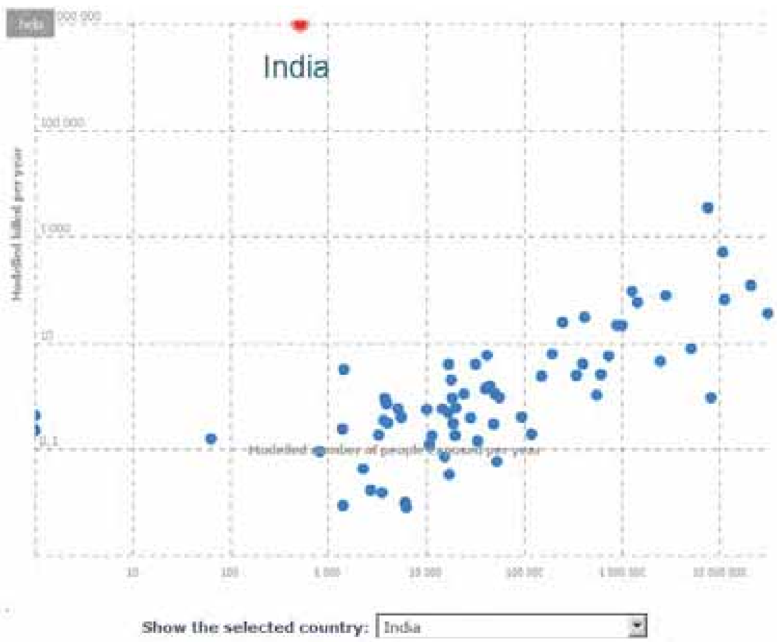


Figure C1.3: Vulnerability for tropical cyclones

Source: http://www.preventionweb.net/english/maps

If you look closely at Figure C1.3 you will see from the x–Axis that about 800 people are exposed per year in India to cyclones and that about 10 million die in India per year due to cyclones. Does this make sense to you? No? Well to us it neither does. However, we hope that this example shows you, that there is a need to compose good graphs using the appropriate statistical techniques for your well thought model. (Eventually the author has just mixed up labelling of the x–and y–axis). In this case, Figure C1.3 is very similar with Figure C1.2. Unfortunately this graph does not give the ratio of persons killed versus exposed but just displays the number of exposed and killed.

1.3 Independent exercise: Cyclones

1. Question: Many People in the world are killed by Cyclones – Where does India stand? Look at the International Comparison or mortality due to Cyclones
2. Many others countries have more people killed? Yes/ No Hint: Look at the introduction of Module A or use your own choices Look at the Figure C1.2 and look what the horizontal legend and the vertical legend say. Do we look for the absolute number or the relative number compared to the total population?
Answers 1. Second 2. No in absolute number.

1.4 Earthquakes

1.4 Scenario 2 – Earthquakes

Figure C1.4 carries a clear message: India is a high risk country regarding earthquakes. Now let's compare Figure C1.4 and Figure C1.5 after we have corrected the incorrect labelling of the axis (As it stands, the authors say that there are about 10000 exposed and 10 million killed. However, this can only be the reverse).

Once we have corrected the labelling we find that the x–axis is the same for Figure C1.4 and Figure C1.5. The difference on the y–axis is that Figure C1.4 is for killed per million inhabitants and Figure C1.5 is for killed per people exposed. So the only difference is the scope of the people it refers to. As mentioned previously comparing the number or persons killed with the number of inhabitants might distort the existing relationship since there might be big differences in the number of inhabitants and the number of persons exposed to the risk.

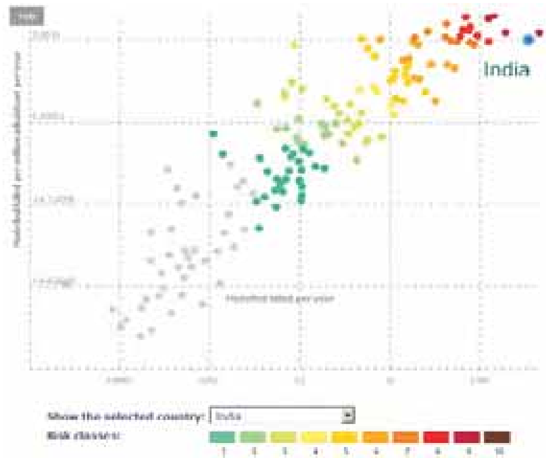


Figure C1.4: Mortality Risk Index (MRI) for earthquakes

Source: http://www.preventionweb.net/english/maps

In our view Figure C1.5 therefore should be preferred to Figure C1.4 but labels of axis have to be reversed.

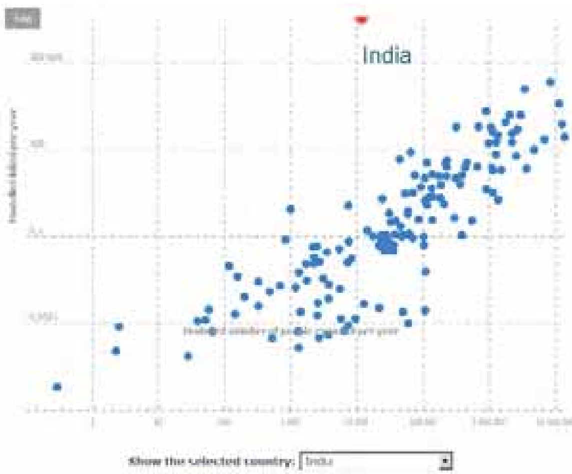


Figure C1.5: Vulnerability for earthquakes

Source: http://www.preventionweb.net/english/maps

1.5 Independent exrcise: Earthquakes[1]

1.	Question: Worldwide, many people are killed by earthquakes – Where does India stand? Look at Figure C1.4: Which of the following statements about earthquakes in India is correct? a. Not only are the most people killed in India compared to other countries, but also the most people per million inhabitants die each year in India. b. India is second in absolute terms of people killed by earthquakes, but relatively several other countries have even more casualties per million inhabitants. Nevertheless India remains in the highest risk class.
2.	Many other countries have more people killed by Earthquakes than India. a. Yes / b. No Feedback wrong (hint): Look at the Figure C1.4 and look what the horizontal legend and the vertical legend say
Answers 1. b 2. b	

[1] The earthquake data include tsunamis

1.6 Scenario 3 – Floods

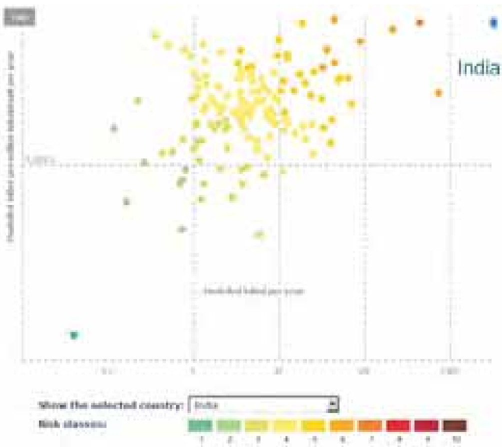


Figure C1.6: Mortality Risk Index (MRI) for floods

Source: http://www.preventionweb.net/english/maps

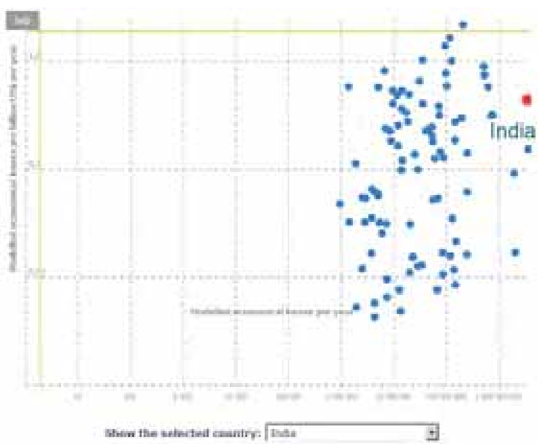


Figure C1.7: Economic Risk Index (ERI) for floods

Source: http://www.preventionweb.net/english/maps

1.7 Independent exercise: Floods

1. Question: Many People in the world are killed by floods– Where does India stand?

   Look at Figures C1.6 and C1.7:

   1. Which of the following statements about floods in India is correct?

   a. Not only are the most people killed in India compared to other countries, also the most people per million die each year in India. In this first rank India shares with few other countries

   b. India is second in absolute terms of people killed by floods, but relatively several other countries have more casualties per million inhabitants by floods than India.

   c. Many other countries have more people killed by floods than India. It is not such a big problem.

2. Floods also cause severe economic problems. Many people lose their housing, crops and essential belongings. Where does India stand in relative terms compared to other countries?

   Look at the International comparison in Figure C1.7, due to floods many people lose their belongings in floods, their crops are destroyed, homes washed away. Which of the following statements about floods in India is correct?

   a. People live in India and economy suffers from the highest absolute losses by floods. Only one other country has to lament as many losses as India.

   b. Many other countries have more economic losses by flood and many countries count more people of being exposed than India concerning floods.

<u>Answers</u>

1. b
2. a

Disaster Scenario in India compared to other countries

India is one of the worst disaster–affected countries in the world and accounts for one fifth of global death count due to floods. Floods occur mostly during southwest monsoon starting from June to September. About 40 million hectares of land in the country is prone to floods as per National Flood Commission. The annual average of land and crop area affected due to flood is about 18.6 million hectares and 3.7 million hectares respectively. But other disasters are not less fearful: The exposure to cyclones is of estimated 7,600,000 Indians, to drought of 59,000,000 to flood of more than 15,000,000, to landslide of 180,000 and to earthquakes of more than 3,000,000.

Source of data: 2009 Global Assessment Report

1.8 Independent exercise

Type of task: Multiple Choices

Sometimes more than one selection is correct, sometimes none.

1. Question: Go to http://www.preventionweb.net/english/maps then go to menu "Graphs" and answer the following questions:

   Which country has the highest economical loss by cyclones?

   a. United States of America

   b. Japan

   c. China

   d. None of these

2. Which country has the highest number of casualties by earthquakes?

   a. China

   b. Japan

   c. Bangladesh

   d. None of these

Answers

1. b)
2. a)

1.9 Conclusion from exercises

The analysis of the Figures in this unit, with all their shortcomings and methodological errors, shows an extremely high risk of casualties and also economic losses due to disasters in India. Physical and environmental conditions in India are very prone to disasters and more information on details, regions and frequencies will be found in the next units.

On the positive side, it can be stated that the relative number of casualties due to disasters is high but not as severe as in the worst affected countries. One conclusion could be that some progress has been made already in India by initiating Disaster Risk Management measures.

Yet, the relatively high international ranking in relative casualties over all major disasters types shows that much work is still necessary in order to reduce the risks for Indian citizens to fall victim of the recurrent disasters and in order to reduce the peril of losing housing and personal belongings.

The further units and modules of this course will embark on tools and methods necessary to help improving the quality and effectiveness of a sustainable Disaster Risk Management for the future.

2

LEARNING UNIT C–2: FREQUENCY AND IMPACTS OF MAJOR DISASTERS IN INDIA

2.1 What is this unit about?

In this unit we will look at risk and frequency[2] of disasters over the last 30 years, as expressed through statistical information. These data are important information as these tell the Disaster Risk Management where the major hazards are, how often disasters occurred and what their impacts were.

In 2007, over 16,000 lives were lost in a total of nearly 1,000 natural catastrophes worldwide. The biggest one was caused by cyclone Sidr (http://en.wikipedia.org /wiki/cyclone_Sidr) which devastated the coastal areas of Bangladesh in mid–November, resulting in 3,300 deaths and leaving more than three million homeless and over 50,000 injured. Already in July and August of the same year, the monsoon had brought extensive flooding to India, Nepal and Bangladesh, causing over 2,000 deaths.

The long–term trends are also clear: recent years have seen a dramatic increase in the frequency and intensity of weather–related natural disasters. With windstorm and flood losses on the advance, weather records worldwide are being shattered. Climate change has dramatic consequences: faster–rising sea levels, retreating glaciers and changes in the seasons.

[2] Frequency is the number of occurrences of a repeating event per unit time. It is also referred to as temporal frequency. The period is the duration of one cycle in a repeating event, so the period is the reciprocal of the frequency.
(Source: http://en.wikipedia.org/wiki/Frequency)

If we examine the consequences of global weather–related disasters, we note that many countries in the early stages of development are particularly affected. The World Bank categorizes countries into groups[3] 1 (rich) to 4 (poor) according to their GNP. The following conclusions can be drawn from the figures for the period 1980–2007:

- Approximately half of the world's 14,500 recorded weather–related natural catastrophes occurred in highly developed countries (G1), a third in groups G3 and G4.
- Over two thirds of the one million deaths caused by disasters are accounted for by the lowest income countries (G4), whereas only 12% relate to high–income countries (G1).
- Whilst, quite unsurprisingly, monetary losses arose for the most part (74%) in G1 and G2 countries, the poorer group G3 and G4 countries bore 26% of the burden. But there is a serious lack of relief or financial cover to mitigate losses in the poorer countries, a mere 1% of insured losses is recorded in group G3 and G4 countries.

Source: Munich Re Foundation &
http://www.preventionweb.net/files/2915_LosterPoorpeoplearetherealloserswеltsichtendossier52008.pdf

2.2 Frequency and impacts of disasters in states of India: A comparison

What is the meaning of occurrences, regular, irregular, repeating and non–repeating patterns? And the impacts of disasters are they related to the repetition or are they not?

We will explain the differences and the impact of Disaster Risk Management on the impacts of earthquake and cyclone disasters. At first we will look at an overall picture of disasters from the following source:

> EM–DAT: The OFDA/CRED International Disaster Database, http://www.emdat.be (Université Catholique de Louvain – Brussels – Belgium).

This data base comprises data on disasters worldwide for as long as historical record keeping goes, in several cases back to 1800. It also comprises both natural as well as technical (man–made) disasters. We will look at the last 30 years of disasters in India.

[3] World Bank Income group: Economies are divided according to 2009 GNI per capita, calculated using the World Bank Atlas method. The groups are: low income, $995 or less; lower middle income, $996 – $3,945; upper middle income, $3,946 – $12,195; and high income, $12,196 or more (all income in PPP $ per year). *Source: http://data.worldbank.org/about/country–classifications and WDI*

The information on natural disasters presented here is taken from EM–DAT: The OFDA/CRED International Disaster Database. In order for a disaster to be entered into the database at least one of the following criteria has to be fulfilled:

- 10 or more people reported killed
- 100 people reported affected
- a call for international assistance
- a declaration of a state of emergency

The information on industrial or technological disasters presented here are following the same criteria to be fulfilled:

Our goal is to identify the highest damages done by disasters, rank them according to frequency and impact and relate natural to industrial disasters in this respect.

Like all data sources, the information has to be taken in with caution. What is written on paper (or electronic devices) is not always true, sometimes not even close to the truth. But if the information is deficient or incomplete, we have to improve the situation: Provide access to better information and better analysis. This is one of the purposes of the course.

In the case of the Centre for Research on the Epidemiology of Disasters (CRED) data, one has to be aware of the fact that there might be underreporting of past events. A disaster of less 10 people killed might not make it into the media and reach CRED. Also the numbers quoted are sometimes differing from database to database. CRED is using official figures wherever possible. However, other databases might use figures from other sources or may adjust official figures because new information has come up and initially published figures have not been updated.

A last point to be mentioned here briefly is the definition of "one event". Sometimes incidents that are somewhat related to each other are grouped together (i.e. all victims to the monsoon floods in one state are grouped together, whereas it might as well be legitimate to see each local flooding as a separate incident. Therefore we have to remember that when analysing data, we have always to ask ourselves:

- What are the criteria used?
- What are the definitions?
- Is the concept applied consistently within the database (i.e. using official government records for one state as a source and in another state using media articles as a source; or grouping of all monsoon floods, but separating the snow avalanches in winter as separate events)?

Guided Exercise

This is an exercise looking at information available from the above mentioned database for India. For many of the results of disasters we have two sets of information for each disaster: the number of occurrences and the average of results per event. The results are people killed, affected and damage in money. We will try to calculate a ranking according to results of frequencies and averages. This is not a course on EXCEL formulas (!) and it should only help to facilitate the ranking. Also discussions might be open to whether the classification of disasters might be modified, but for this simplified example for classification we should postpone the discussion:

Load EXCEL File “Natural DisastersIndia_Last 30years.xls” and “IndustrialDisastersIndia_Last 30years.xls” and find the five top scorers for the Natural Disasters

(Frequency, People Killed, Affected and Damages associated with each disaster

Hint: Try to use EXCEL or Open Office formula (this is for the Natural Disaster File):

(=+IF (C3<>"–";(+IF(C3<>"";RANK(C3;C$3:C$46);""));"") and copy this formula to get relative (column wise) ranks for each of the empty columns. You can also sort the data according to these numbers if that seems easier for you.

Try to use relative formatting, (refer to EXCEL help if necessary) to show first 5 ranked disasters for Natural disasters and first 3 ranked disasters for Industrial disasters.

The result should look somewhat like this (this is of course a sample of the solution only).

Table C2.1: Impact of disasters

		Frequency	Impact–kille	Impact–affecte	Impact–damag
	Heat Wave	9			
	ave.per event		7	32	20
Flood	Unspecified	3	18	35	14
	ave.per event		6	2	2
	General flood	1	20	9	11
	ave.per event		2	3	1
	ave.per event		27	11	12
	Landslide	5	30	31	15
	ave.per event		12	10	27
Storm	Unspecified	8	36	25	23
	ave.per event		14	18	28
	Local storm	9	34	26	18
	ave.per event		13	17	5
	Tropical		33	24	19
	cyclone	2			
	ave.per event		3	5	3

The Table C2.1 shows the ranking of cyclones as second in frequency, third according to number of people killed and damage affected and ranked fifth according to people affected by this type of natural disasters compared to others. In comparison flood ranks first in frequency and damage, second in people killed and third in people affected.

What to learn from the Exercise?

The analysis of the files Solution_NatDisastersIndiaDetail30years.xls and Solution_IndDisastersIndDetail30years.xls (in CD) show clearly some deficits:

The number of disasters is certainly not complete, the only Tsunami listed is the most recent one on 26–Dec–2004, no others are listed. However, tsunamis are not unique and isolated events but closely related to earthquakes and costal surges and floods and could be regarded in combination with these.

If we look at the man–made disasters we find that another doubtful figure is the one on transport disasters because there were certainly more accidents in the last 30 years than the ones listed in the EM–DAT database by CRED. However, taking into account the above given definition of disasters (i.e. more than 10 killed), the figures of EM–DAT might make more sense. Especially road accidents might claim smaller number of casualties than 10 and all other accidents do not enter the data base.

Comparing this to another national source: "The National Crime Records Bureau, Delhi", will give much higher and comprehensive figures. This source shows much higher figures than CRED. It shows here a strong discrepancy to the CRED source: Fatal casualties vary among the 1 lakh figures throughout the last years. Therefore we can assume that the CRED source is showing not very realistic data compared to the National Crime Records Bureau. This shows that before dealing with data from various sources, it will be necessary to validate the data in order to ensure their usefulness before any statistical analysis is taken in hand.

Table C2.2: Road traffic fatalities in India

Year	Fatalities	Population (million)	Fatalities/million persons
1997	77,000	955	81
1998	79,000	971	82
1999	82,000	987	83
2000	78,900	1,002	79
2001	80,900	1,027	79
2002	84,059	1,051	80
2003	84,430	1,068	79
2004	91,376	1,086	84
2005	98,254	1,103	89
2006	105,725	1,120	94
2007	114,590	1,136	101

Source: National Crime Records Bureau, Delhi

Comparing the two data files (Solution_NatDisastersIndiaDetail30years.xls and Solution_IndDisastersIndDetail30years.xls). Whereas the frequency is the figure where accidents of all kinds will exceed the frequency of natural disasters, all other damages inflicted by natural disasters exceed by magnitude those of industrial disasters.

2.3 Conclusion from this unit

The risk of industrial disaster might increase with progress and industrialization. Sticking to the current results the comparison shows us something similar to the summed up graphics of the first module:

Floods are front–runners in all categories (i.e. people killed):

Only earthquakes have higher frequency and the higher damages, in number of people killed compared to floods and floods rank third for the number of people affected by this disaster compared to others.

Earthquakes claimed the highest number of victims and it ranks fourth as of damage–impact as compared to other disasters.

Drought affects the highest number of people but claims relatively less damages and lives.

Cyclones are also front–runners in all categories, second in frequency, third in casualties and damages (but might rank second if other storm damages are added) and fifth in rank of people affected.

Other disasters like **Landslides** or **Epidemics** occur quite often (fifth and fourth) but claim in total less victims and damages.

As a conclusion of the exercise we can state, if we want to learn more about Disaster Risk Management it might make sense to focus on the big four (natural) disasters, not ignoring other important risks and dangers like Landslides, Epidemics and industrial and technological hazards, particularly in terms of newly emerging risks caused by Climate Change . However, our course length and scope asks for concentration and reduction.

This goes very much in line with the **International Strategy for Disaster Reduction** Report of Jan/2010 (Figure C2.1):

"Earthquakes are the deadliest natural hazard of the past ten years and remain a serious threat for millions of people worldwide as eight out of the ten most populous cities in the world are on earthquake fault–lines," ...

"Disaster risk reduction is an indispensable investment for each earthquake–prone city and each community. Seismic risk is a permanent risk and cannot be ignored. Earthquakes can happen anywhere at any time. Risk reduction will be a main priority in the Haiti reconstruction process, and we will be working with our partners to ensure that it is central in the reconstruction" continued Margareta Wahlström, Secretary–General for Disaster Risk Reduction of the United Nations."

Figure C2.1: Time trend of reported natural disasters, 1975–2009

Sources: Centre for Research on Epidemiology of Disasters (CRED)

"3,852 disasters killed more than 780,000 people over the past ten years, affected more than two billion others and cost a minimum of 960 billion US$."

"In terms of human losses, Asia is the continent that has been struck again and again by disasters during the last decade, accounting for 85 per cent of all fatalities."

"After earthquakes, storms (22%) and extreme temperatures (11%) were the most deadly disasters between 2000 and 2009. The most deadly disasters of the 2000 decade were the Indian Ocean Tsunami, which hit several countries in Asia (2004) leaving 226,408 dead; Cyclone Nargis, which killed 138,366 people in Myanmar (2008); and the Sichuan earthquake in China (2008), causing the deaths of 87,476 people. 73,338 people were also killed in the earthquake in Pakistan (2005) and 72,210 in heat waves in Europe (2003)."

"The number of catastrophic events has more than doubled since the 1980–1989 decade. In contrast, the numbers of affected people have increased at a slower rate."

Sources: Centre for Research on Epidemiology of Disasters (CRED)

3

LEARNING UNIT C–3: SELECTED AREAS OF DISASTERS IN INDIA AND THEIR FREQUENCY AND IMPACTS

3.1 What is this unit about?

In this unit we will look at the risk and frequency of past disasters under regional aspects, as expressed by statistical information. This is important information as it tells about the past disasters in states under statistical aspects. Which disasters had occurred at the highest risks or outside the risky periods, which ones affected people the most and which caused the most economical damage and whose effects were probably mitigated by preparation and foresight. This module will hint at regions with best practices to cope with disasters in the past. Here we will again consult available statistical information on prevention, mitigation, response and recovery in selected areas.

The approach in this module will be to look at the fact based information sources in risk prone states. Since it will be difficult to take all 33 Indian states into consideration, three risk prone states will be selected and the risks in these states will be assessed in some detail and by regional districts without being able to go too much into detail. More important still will be a historical view on action and DRM processes in these states during and past disasters.

For this purpose we will rely on the exercise from previous units to identify risks by state. This scheme and the method used have certainly room for improvement (e.g. there is no weighting involved, classifying landslides risk the same as earthquake risks) but at least it provides an indicator of natural hazard risks by state for the most dangerous natural disasters. All we will have to do is to apply some ranking to identify more disaster prone areas against others.

3.2 Guided and independent exercise

To identify risk prone states we will refer to our exercise from Module B /Unit2 and load the solution file "RisksDisastersStatesSolution.xls" and find the top five scoring states or territories for being affected by cyclones, earthquakes, tsunami/storm surge, floods, landslides and drought.

Hint: Try to use EXCEL or OpenOffice formula:

+IF(J2<>"–";(+IF(J2<>"";RANK(J2;J$2:J$46);""));"")and copy this formula to get relative (column wise) ranks for each of the empty columns.

Try to use relative formatting, to show first 5 ranked disasters for Natural disasters and first 3 ranked disasters for Industrial disasters.

As a result of the exercise we will come up with something like this:

Table C3.1: Ranking of disasters in different states of India

State	Cyclone	Earthquake	Surge/Tsunami	Flood	Landslide	Drought	Average	Rank
Andhra Pradesh	5	3	5	5	1	5	4,0	2
Gujarat	5	5	5	5		5	4,2	1
Orissa	5	2	5	5		5	3,7	3
Tamil Nadu	5	2	5	4		5	3,5	4
West Bengal	5	4	5	5			3,2	5

The Table C3.1 does not mean that other states and territories do not suffer from devastating disasters but for the moment we have identified the 5 states ranked first (the territories ended up with lower ranks) and for sure it's questionable to weight all the disasters equally. But like all our statistical exercises you will have to look for the technical learning aspect ("how does it work").

The examples in this Module C should help to clarify a historical compendium of disasters and highlight the use of fact based information. The purpose again is to make participants aware of the possibilities of information available, also of the misuse of information and provide hints how to use them best.

3.3 History of natural disasters in Andhra Pradesh

The history of disasters in Andhra Pradesh is stated very briskly on the Disaster Risk Management site of the Government of AP – http://disastermanagement.ap.gov.in/website/history.htm. This history of natural disasters in Andhra Pradesh allows an overview, however it is not a statistical compendium.

Andhra Pradesh is exposed to cyclones, storm surges, floods and droughts. A moderate to severe intensity cyclone can be expected to make landfall every two to three years. About 44 % of the state is vulnerable to tropical storms and related hazards. In India, the cyclones develop in the pre–monsoon (April to May) and post–monsoon seasons (October to December), but most of them tend to form in the month of November. Cyclones on the east coast originate in the Bay of Bengal, the Andaman Sea or the South China Sea, and usually reach the coastline of Tamil Nadu, Andhra Pradesh, Orissa and West Bengal, which are the most vulnerable to these types of hazards. Two of the deadliest cyclones of the last century, with fatalities of about 10,000 people in each case, took place in Orissa and Andhra Pradesh during October 1971 and November 1977 respectively. The super cyclone of Orissa in 1999 caused large scale damage to life and property.

The states bordering the Arabian Sea on the west coast are not completely safe either, as Kerala, Gujarat – and to a lesser extent Maharashtra – are also prone to cyclones. With a frequency of four cyclones per year, one of which usually becomes severe, the Bay of Bengal accounts for seven percent of the annual tropical cyclone activity worldwide. The level of human and property losses that cyclones cause around the Bay is very high. Once the cyclones enter the mainland, they give way to heavy rains which often translate into floods, as it was the case with the damaging cyclone–induced floods in the Godavari delta, in August of 1986.

Many drought prone areas adjacent to coastal districts in eastern maritime states are thus vulnerable to flash floods originated by the torrential rains induced by the cyclonic depression. In addition to cyclones and its related hazards, monsoon depressions over the north and central areas of the Bay of Bengal move until reaching north and central India, including portions of Andhra Pradesh, bringing heavy to very heavy rains and causing floods in the inland rivers between June and September.

In Andhra traditionally, the flood problem had been confined to the flooding of rivers. But the drainage problem in the coastal delta zones has worsened particularly in recent times, multiplying the destruction potential of cyclones and increasing flood hazards. A critical factor is maintenance of irrigation systems. On several occasions, deaths have been caused by breaches in canals as well as over–flooding caused by silting and growth of weeds in irrigation canals.

Impact of Disasters in AP

The regular occurrence of disasters both natural and manmade in coastal Andhra caused a series of repercussions on the state's economy, its development policies and political equilibrium and daily life of millions of Indian citizens?

Andhra Pradesh is battered by every kind of natural disaster: cyclones, floods, earthquakes and drought. The coastal region suffers repeated cyclones and floods. The 1977 cyclone and tidal wave, which resulted in great loss of life, attracted the attention of the central and state Governments of India and the international donor communities, as did those of 1979, 1990 and 1996. The floods in the Godavari and Krishna Rivers caused havoc in the East and West Godavari and Krishna districts.

Earthquakes in the recent past have occurred along and off the Andhra Pradesh coast and in regions in the Godavari river valley. Mild tremors have also hit the capital city of Hyderabad, for example in September 2000.

Social and economic life of AP's population is characterized by recurring natural disasters. The state is exposed to cyclones, storm surges, floods, and droughts. According to the available disaster inventories, AP is the state that has suffered the most from the adverse effects of severe cyclones.

More than sixty cyclones have affected AP this century. The incidence of cyclones seems to have increased in the past decades, to the extent that severe cyclones have become a common event occurring every two to three years, repeatedly and severely affecting the state's economy while challenging its financial and institutional resources. Almost 9 million people are vulnerable to cyclones and their effects in Coastal AP, 3.3 million of those who belong to communities located within five km of the seashore. The deadliest cyclone in the last twenty years took place in November 1977 killing about 10,000 people. More recently, the May 1990 cyclone, with a death toll close to 1,000 people, caused about US$1.25 billion damage in ten districts, including the entire coast. Between 1977 and 1992, about 13,000 lives and 338,000 cattle were lost due to cyclones and floods, and nearly 3.3 million houses damaged.

3.4 Guided exercise on cyclones in Andhra Pradesh

This exercise gives a comprehensive briefing of the disaster situation of AP also in numbers (however, they are not very systematic). They also refer to the situation of other states in India are especially related to cyclones and their devastations. The source http://disastermanagement.ap.gov.in/website/history.htm reveals a table referring to cyclone damages and occurrences.

Load and look at the simplified graph of the results of cyclones in AP: “Cyclones Damages and Impacts in AP.xls”. The graph looks like this: The **bars** display people affected in Numbers (multiplication of affected population to make graphs magnitude comparable). The **lines** display damages in hectares of crop area and Rupees (Multiplications has been added to make graphs magnitude comparable).

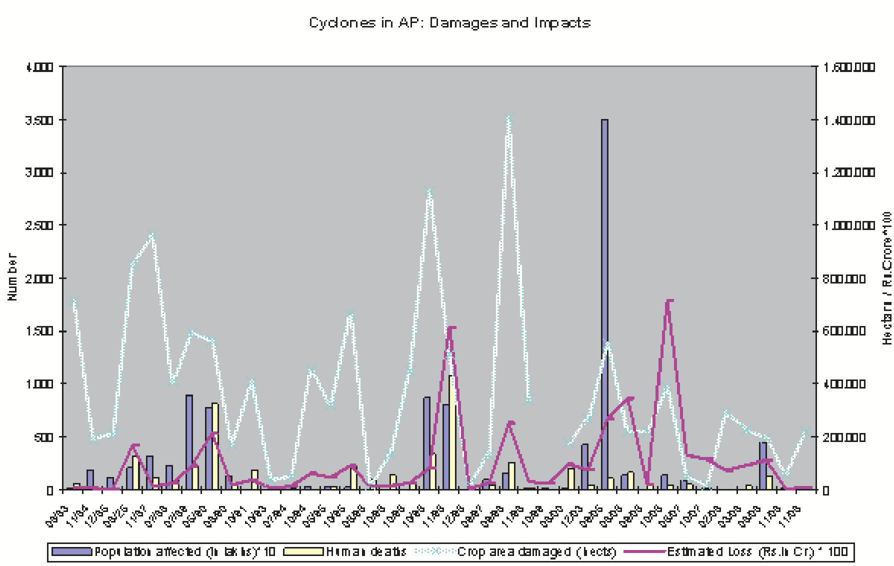


Figure C3.1: Cyclones in AP – damages and impacts

Look at the Solution of Cycles Damages in AP (Figure C3.1 and file “Solutions–Cyclones Damages and Impacts in– AP.xls” sheet: Casualties) and look at the Graphs showing linear trends (trend lines) over the last 30 years (A red trend line) Trend lines in EXCEL are added simply to a graph clicking on the series and choosing menu “Add trend line”. How they are calculated will be explained in the next module.

Please answer the following questions:

1. What can you say about the frequency of cyclones in AP over the last 30 years? (Yes/No answers)

   a) The frequency increased

   b) The frequency remained the same

   c) The frequency decreased

2. What can you find about people killed by cyclones in AP over the last 30 years?

   a) The number of people killed increased.

   b) The number of human life loss is decreased.

   c) The number of people killed remained the same

3. What can you find about people affected by cyclones in AP over the last 30 years?

   a) The number of people affected by cyclones increased

   b) The number of people affected by cyclones decreased

   c) The number of people affected by cyclones remained the same

4. What can you find about crop areas damaged by cyclones in AP over the last 30 years?

   a) The crop areas damaged by cyclones increased

   b) The crop areas damaged by cyclones decreased

   c) The crop areas damaged by cyclones remained the same

5. What can you find about economic losses by cyclones in AP over the last 30 years?

   a) The economic losses by cyclones increased

   b) The economic losses by cyclones decreased

   c) The economic losses by cyclones remained the same

Answers

1. a)

2. b)

3. a)

4. a)

5. a)

3.5 Independent exercise: Disasters in Orissa and Tamil Nadu

There are comparable data for Orissa and Tamil Nadu but you may see the differences yourself

Sources: http://www.desinventar.net/DesInventar/index.jsp

Look at the Web Site and open the EXCEL File: "Orissa 1970–2006 + Tamil Nadu 1976–2006.xls" being generated from the data from this file. It contains three graphs for each state: People killed by disasters, Houses damaged and destroyed by disasters and People affected by disasters. Note that the graphs in the Web Site in some instances do not reflect the data on the same site.

For all three graphs trend lines have been added.

Look at the graphics and results

1. Please answer the following questions for both states: (3 Yes/No answers) + 3 open question

   a) You can distinguish and identify cyclones affects separated from the impact of other disaster.

   b) The trend line shows an increase of disasters and their effects for both states.

   c) The number of houses destroyed is often higher than the number of houses damaged.

2. Please answer the following questions:

Can you use these statistics for any Disaster Risk Management activities and how?

(Hint: example Identification and mapping of hotspots, understanding trends, evaluation, validating models as so on)

- If you compare the AP statistics on cyclones with the disaster statistics for OR and TN. Which are the strong points of the OR and TN statistics?
- If you compare the AP statistics on cyclones with the disaster statistics for OR and TN. Which are the weak points of the OR and TN statistics?

Answers

1. b)

3.6 Andhra Pradesh: Conclusion from exercise

If we can believe the figures from the exercise "Cyclones in Andhra Pradesh", it can be concluded that the Cyclones Damages and Impacts in AP.xls shows **positive results** as far as Disaster Risk Management is concerned. The reasoning is that even though the number of people affected increased, the number of human life loss is decreased.

The same is true for crop damages. These decreased over the last 30 years. However, the economic losses increased but this is certainly reflecting inflation figures (the monetary data have almost certainly not been inflation adjusted!). Moreover the data do not indicate if rising economic losses are due to development and increasing costs of infrastructure. Even though or reasoning is rather superficial, the information here is **rather uncertain.**

As a **hypothesis** based on these data one can deduct that the effect of Disaster Risk Management in AP – at least as far as the imminent cyclone threat is concerned – has been successful. Based on this fact based information, and in absence of further data on Disaster Risk Management, this conclusion can only be a first assumption. The increase in frequency of cyclones is certainly alarming and may be a result of climate change.

The web site of the Disaster Risk Management site of the Government of AP (http://disastermanagement.ap.gov.in/index.htm) contains comprehensive material about:

1. History of disasters in AP
2. Natural disasters in AP:

- Floods (excellent mapping and historical data)
- Cyclone (mapping and historical data as seen above)
- Drought (excellent mapping and historical data)
- Tsunami (excellent mapping and historical data)
- Earthquake (only 3 earthquakes in last 100 years)

3 An insurance scheme promoted by the Government of AP

4. Information about national and international agencies like:

- NDMA and others
- Manuals
- Mock Drill
- Preventive Measures
- Weather Forecast from IMD like Rainfall

As a first and still rather superficial assumption: (It seems that Andhra Pradesh is prepared in terms of Disaster Risk Management as far as the most threatening hazard – cyclones – is concerned. It seems that the use of fact based information by the Government of AP is pertinent.

3.7 Gujarat: History of disasters

The 2001 Gujarat earthquake occurred on January 26, 2001, at 08:46 AM. The location of the epicentre was Bhuj, Gujarat. With a moment magnitude (Mw) of between 7.6 and 8.1, the quake killed more than 20,000 people and injured another 167,000 and destroyed nearly 400,000 homes.

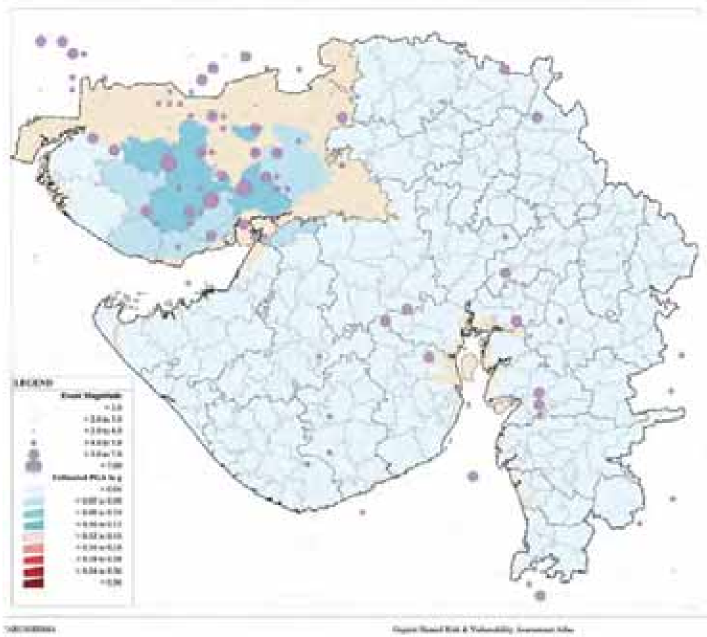


Figure C3.2: Gujarat earthquake hazard risk zonation: 25 years return period estimated mean taluka peak ground acceleration (PGA) (in g). 1 Gal (g) is equal to 0.01 m/s²

Source: http://en.wikipedia.org/wiki/Gujarat_earthquake_of_2001

The GIDM (Gujarat Institute of Disaster Management) conducts hazard zonation mapping, showing maps of earthquakes, cyclones, floods, and tsunami and storm surge.

Source: http://www.gsdma.org/hazard/eq/25_year.jpg

The Earthquake mapping (in Figure C3.2 shown for a period of 25 years) shows a clear peak of earthquake risks in the northwest of Gujarat, the Kutch area (near Bhuj) close to the Pakistani border. Maps with longer periodicity (50, 100, 200 years) confirm this information. No information on damages or impacts of past disasters is available on this page.

We will try to look for more **detailed information.**

3.7.1 Gujarat: More on earthquakes

Let's first take in more general information like a map of earthquakes having occurred in the region:

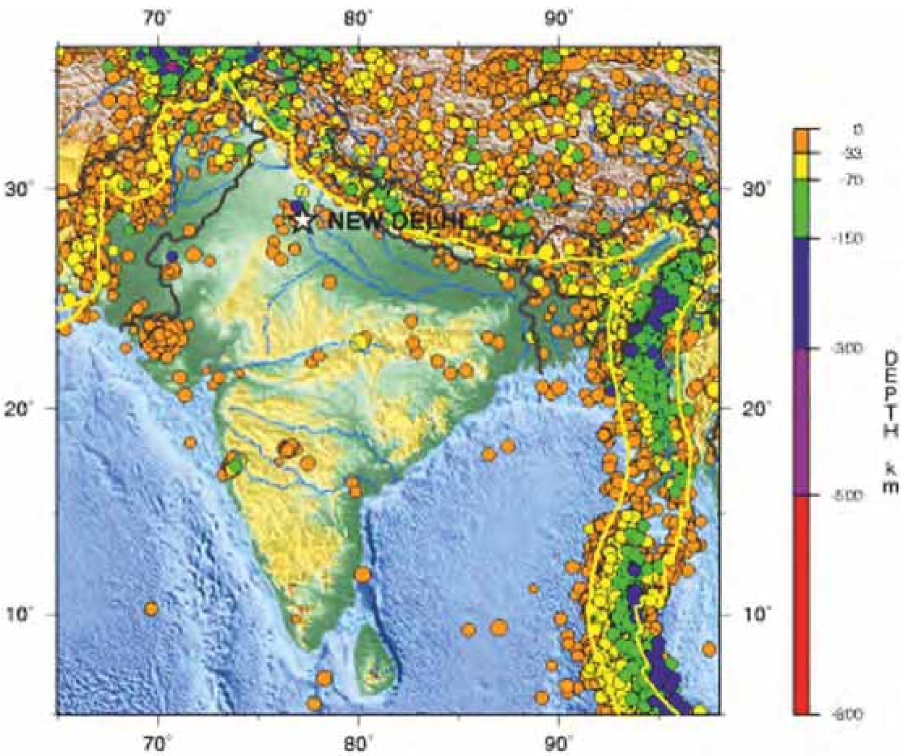


Figure C3.3: Seismicity of India, 1990–2006

Source: http://earthquake.usgs.gov/earthquakes/world/india/seismicity.php

The map in Figure C3.3 shows a high peak of occurrences and frequency of earthquakes in Gujarat compared to any other region in India. The last "big" earthquakes in India since 1990 confirm this picture (Table C3.2 is showing magnitude as the preceding number):

Table C3.2: “Big” earthquakes in India since 1990

Nr.	Date	Area affected	Loss event	Deaths	Economic losses[4]
1	08.10.2005	India	Earthquake	1,300	no data
2	26.12.2004	India	Earthquakes, tsunami	16,300	est. US$ 2500m
3	26.01.2001	India W, Gujarat. Pakistan	Earthquake	14,970	> US$ 4500 m
4	29.03.1999	India Uttar Pradesh, Chamoli	Earthquake	110	est. US$ 2 m
5	30.09.1993	India Maharashtra, Khillari, Latur, Umbarga	Earthquake	7,601	est. US$ 280 m
6	20.10.1991	India Uttar Pradesh, Almora	Earthquake	2,000	est. US$ 100 m
7	21.08.1988	India	Earthquake	1,000	est. US$ 70 m

Source: http://mrnathan.munichre.com/

Damages and impacts in detail

A comprehensive analysis of damages and preventions mainly about house construction issues is found at: http://www.becs.ac.in/SDME/SkmDarjEQReport.pdf

Certainly different states will have differing demands for sustainable and earthquake ready construction due to local conditions. Another very comprehensive lesson learnt from the Gujarat earthquake is the “Community recovery and its sustainability: Lessons from Gujarat earthquake of India” accessible through:

http://www.ema.gov.au/www/emaweb/rwpattach.nsf/VAP/(99292794923AE8E7CBABC6FB71541EE1)~Community+Recovery.pdf/$file/Community+Recovery.pdf.

Not to forget the generic measures proposed by NDMA like the “Earthquake emergency survival guide”:
http://ndmindia.nic.in/Pocket%20Guide.jpg

[4] Losses are not only by state but encompass all the losses even across national borders.

As a conclusion of the example we can state:

Earthquakes, at least the most destructive ones, occur in intervals largely spaced. This makes them very hard to predict contrary to disasters based on natural repetition cycles like floods and droughts. We will come to earthquake risks in a more comprehensive exercise in the Module D.

Damages and casualties can be extremely high and their numbers amount to about half of all victims of natural disasters.

The unpredictable characteristics of earthquakes are also reflected in the various advises about general earthquake preparedness, especially in areas known to have frequent or large quakes.

Most important are locally adapted contingency plans for the different phases of disasters. A good example is found in the AP Disaster Risk Management although not for earthquakes as they are not of primary concern of AP state authorities.

3.8 West Bengal and Bihar: History of natural disasters

The history of disasters in Bihar is stated very briefly on the Disaster Management Department site of the Government of Bihar:

> "As Bihar is vulnerable to different kinds of disaster due to its geographical and topographical location, it is prone to Flood, Drought, Fire and Earthquake. According to seismic zoning, some parts of the State are in Zone–IV & Zone–V, which can cause devastation as bad as BHUJ (Gujarat) and LATUR (Maharashtra).Disaster Risk Management Programme, which is working under the Disaster Management Department, Govt. of Bihar, is the process of setting up District Control Rooms with the latest art of the technology having computers, fax, scanners, copiers and Internet connectivity so as to develop the data bank related to disasters, consequences and relief management for generating the knowledge base for the people of Bihar. An important part of this mission is providing assistance to prepare for and recover from natural or man–made disasters that can result in great human and economic losses." Source: www.bsdma.org.

No such on–line information is available for West Bengal but a comprehensive article on "Disaster mitigation and management for West Bengal, India – An appraisal" can be downloaded from:

http://www.ias.ac.in/currsci/apr102008/858.pdf.

It states amongst others: Approximately 55.8% of the region is susceptible to floods in WB. Furthermore, complicacy is implicated by the origination of major flood–producing rivers beyond the state jurisdictional limits. Historical records of large floods in the state will be shown. An outline of flood management 'A Monograph on Flood Management' prepared on the basis of hands–on experience of the State Government officials in 2006 recommends a standard operating procedure and will be cited later on. Three phases of actions are specified: pre–flood, during flood and post–flood. The pre–flood phase activities consist of preparatory measures, which involve vulnerability assessment, personnel and organizational database development, viable emergency action plan such as deployment of early warning system, training of personnel for rescue and evacuation, verification and updating of existing search, rescue and evacuation plans, and inventories of essential commodities and relief materials.

Since both states are affected heavily by floods, this type of hazard will be the main focal point of this section.

3.8.1 West Bengal and Bihar: Details on floods

Table C3.3: 10 year record of West Bengal floods

02/08/2000–01/10/2000	Besides flash floods triggered by incessant torrential storms, disaster is also accredited to the opening of sluice gates of dams. The fatalities counted to the tune of 1262, besides affecting millions of people
31/07/2001–01/09/2001	Monsoonal rains caused flooding in Kolkata
21/06/2002–28/08/2002	Flooding in Jalpaiguri, Cooch Behar and Jalpaiguri in north Bengal due to monsoonal rains. Flash floods swamped ten villages, causing four deaths and 11,000 displacements
11/06/2003–10/10/2003	Monsoonal rains caused floods affecting the regions of Darjeeling, Jalpaiguri, Malda and Murshidabad
20.06.2004–07.10.2004	Heavy monsoonal rains affected several districts.
21.–28.10.2005	Heavy rains caused floods in many areas. About 3000 coastal villages were inundated and 60,000 huts and many roads washed away.
07.–27.07.2005	Heavy monsoon rains triggered flash floods and landslides.
24.06.2006–03.08.2006	The regions of Birbhum, Burdwan and Murshidabad were affected mainly from continuous monsoonal downpour.

18.09.2006–05.10.2006	Monsoonal rains and tropical cyclone–driven storms in the Bay of Bengal hit India and Bangladesh. West Bengal recorded 50 deaths, 300 were injured and 30,000 mud houses destroyed. Heavy rains left large parts of Kolkata city under water; subsequently 2000 people were evacuated from the city.
03.07.2007–22.09.2007	The hazard affected Kolkata and several other districts. Eighty–three deaths were reported, and millions of people were marooned in 3000 villages in coastal areas of the state.
22.09.2007–08.10.2007	Heavy rain from tropical depression in the Bay of Bengal caused flooding leading to 51 deaths, and affecting 3.2 million people.

Source: http://www.iagwestbengal.org.in/IAG/Activities_files/History%20of%20Disasters/Floods.pdf

What is noticeable here is that historical information on floods is rather unspecific (at least from this source), information is not detailed enough (which river, excess degree of river gauge etc.) to deduct future measures.

Also damages would be useful to be collected on most decentralized level to react to future disasters.

3.8.2 Guided exercise: Floods in Bihar

Look at the “DataSet_FloodBihar.xls” and create graphs showing linear trends of flood disasters in Bihar: Human casualties, property damages in money terms, crop damages in area and money terms. In the last graph compare the two trends for crop area damaged and value of crops damaged.

If you cannot do the graphs yourself refer to the tutorial or proceed with Exercise 2.

3.8.3 Independent exercise: Floods in Bihar

Looking at the “Solution_FloodBihar.xls” of the data set and the graphs created from it – Please answer the following questions:

1. What can you say about the casualties due to floods in Bihar over the last 20 years?
   a) The casualties increased (True)
   b) The casualties remained the same (False)
   c) The casualties decreased (False)

2. What can you say about property damages due to floods in Bihar over the last 20 years? a) The property damages increased (False) b) The property damages decreased (True) c) The property damages remained the same (False)
3. What can you find about crop damages in area of land and financial losses to floods in Bihar over the last 20 years? a) Both figures increased (True) b) Both figures decreased (False) c) The land area losses had weaker increases than the financial losses (True) d) The land area losses had stronger increases than the financial losses (False)

3.8.4 West Bengal and Bihar: Conclusions on flood disasters

It looks that the floods in both states are frequent incidents with ever increasing number of casualties and damages and economic losses. There might be a reason for this increase regarding population growth and inflation biased monetary figures. One reason might be the lack of public capability to cope with disasters which are almost predictable by location, month and week.

Certainly some of the floods (like the one in 2004) had unprecedented magnitude and settling schemes and other precarious and harmful conditions may have increased the potential danger. But measures are possible to warn and mitigate hazards damages. We will deal with this in the next modules and we will conclude with a remark from 08.09.2008 from the "Time online".

> **"India's Floods: A manmade disaster?**
>
> After the river breached, it headed south and soon flooded the villages in its path. More than 2.1 million people in the worst–hit parts of Bihar are not only homeless but stranded, and 55 have been killed as the floods washed out the roads and railroad lines that connected residents to the rest of the country.
>
> The pictures of those stranded — standing on bits of highway but surrounded by water and looking up at the helicopters they hope might save them — will look familiar to anyone who saw

the images of the rooftops of New Orleans during Hurricane Katrina. India has grown accustomed to natural disaster, but like its North American cousin, this one also looks manmade."

Read more:
http://www.time.com/time/world/article/0,8599,1837449,00.html#ixzz0khWpOUZi

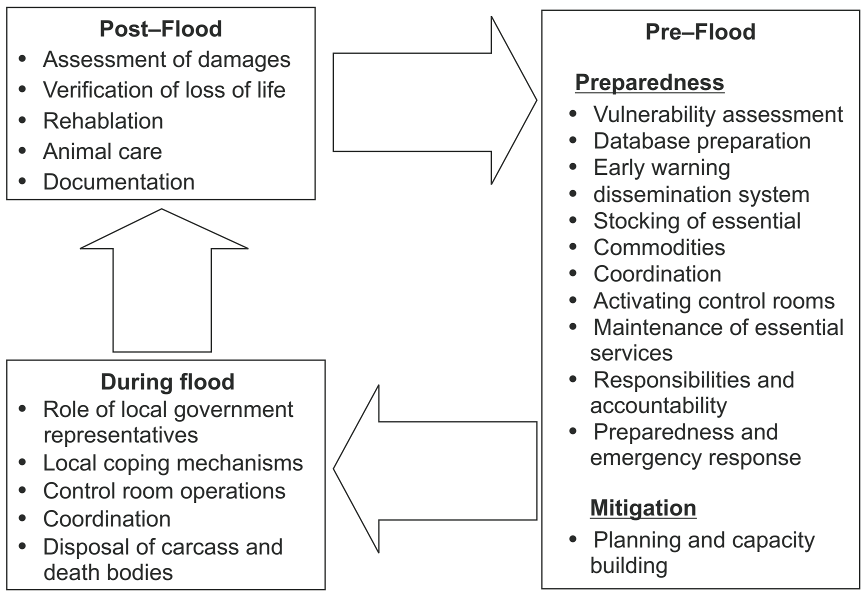


Figure C3.4: Standard operation procedure for managing flood hazards

The conclusion so far should be that information is abundant and due to excellent services of metrological data, remote sensing and mapping by satellite services (especially NRSC: http://dsc.nrsc.gov.in:14000/DSC/index.jsp) to assess dangers and post–disaster hazards diligent use of fact based information. Better information should bring improvement to all phases of Disaster Risk Management especially concerning floods in West Bengal and Bihar.

The encouragement of this course should not only be to use and analyse existing data but also to foster the production of regular reports on damages and losses. Without systematic collection of this information no improvement can be expected but likewise without use and diligent presentation of disaster information repetition of ever the same is imminent. Even if the cycle of life and death continues forever, men are made to improve their lives, make it more secure, and provide for the unforeseen. The consciousness of prevention is one of the distinguishing features of human existence.

4

SELF–ASSESSMENT MODULE C

Question 1:

Which disaster claims the highest number of human causalities in India in the past 20 years?

Select the correct answers.

a) Earthquakes
b) Floods
c) Cyclones
d) Epidemics
e) Road accidents

Question 2:

Economic damages by disasters

Which are the disasters which claimed the highest economic damages in India in the past 20 years?

Select the correct answers.

a) Floods
b) Earthquakes
c) Cyclones
d) Epidemics

Question 3:

In which of the following disasters India had the highest causalities compared to other countries?

Select the correct answers.

a) Floods
b) Earthquakes
c) Cyclones
d) Epidemics

Question 4:

To which of the following disasters India showed the highest vulnerability compared to other countries?

Select the correct answers.

a) Floods
b) Earthquakes
c) Cyclones
d) Epidemics

Question 5:

What kind of Information the earthquake hazard maps of BMTPC contain?

Select the correct answers.

a) Volcanoes
b) Damage risk zones
c) Elevation
d) Rivers and Dams
e) Cities and Population density

Question 6:

Flood prevention by studying maps and census:

Which of the following remarks are correct?

Select the correct answers.

a) The information of upstream flood risks is as important as meteorological information to mitigate effects of flood disaster.
b) Floods occur without prior notice.
c) Damages by floods in urban and rural areas are very similar.
d) The more detailed the map is, the more useful it can be for Disaster Risk Management.
e) It is easy and quite simple to compare disaster risks among Indian States and Territories.

Question 7:

Comparing the impact of Disasters:

If you want to compare the impact of disasters: Which of the following are the criteria to use?

Select the correct answers.

a) No. of districts and population affected
b) Human deaths & live–stock loss
c) Houses & Crop area damaged
d) Crop area damaged
e) Damage of infrastructure
f) Production loss of industry
g) Amount of international and national financial support
h) Response times to disasters

Question 8:

Comparability of data for DRM:

Which statements are correct?

Select the correct answers.

a) It is better to use a moderately reliable data source than source than none.

b) Definition of vulnerability, damage and exposure are not self-explaining and crucial for comparison.

c) More often than never distorted statistics are produced for a purpose.

d) The more sources you compare the better the chances are for better Disaster Risk Management.

e) As often as intentional deviation from the truth, lack of knowledge and limited expertise is the reason for bad fact based information.

Question 9:

Loking at Disaster Risk Management reports from Indian States?

Which Statements are correct?

Select the correct answers.

a) The percentage of deaths caused by cyclones compared to those caused by other disaster will remain the same throughout the years in one state.

b) Definition of Vulnerability, Damage and Exposure are not self-explaining and will change from state to state.

c) The more precise the types of Disasters and their impacts can be distinguished, the easier it is to measure and evaluate their impacts

d) Publishing no information about the Disaster risk management in a state means that there are no problems related to Disaster Risk Management.

e) A Disaster Risk Management report should not contain tables and graphs but needs elaborate and understandable comments.

Question 10:

Which of the following are manmade disasters?

Select the correct answers.

- Earthquakes
- Floods and their impacts, if taking into account lacking preventions
- Industrial disasters like nuclear plants damage
- Cyclones
- Tsunami.
- Road accidents.

Answers

1. e
2. a & b
3. a & b
4. a
5. b
6. a & d
7. a,b,c, d & e
8. b,c & e
9. b,c & e
10. a, d & e

5 RESOURCES

Mandatory Tools, Media and Learning material to use

- Orissa and Tamil Nadu disaster database from http://www.desinventar.net/DesInventar/results.jsp.
- The national portal of India http://www.india.gov.in/.
- The States and Union Territories http://india.gov.in/knowindia/state_uts.php.
- ISRO – Disaster Management Support Programme Decision Support Centre http://www.nrsc.gov.in/.
- State Domestic Products for calculation of material losses http://mospi.nic.in/rept%20_%20pubn/ftest.asp?rept_id=nad03_1999_2000&type=NSSO.
- Population Data of India, Census of India http://www.censusindia.gov.in/.
- Meteorological Indicators: India Meteorological Department http://www.imd.gov.in/.
- Maps of India http://www.mapsofindia.org/.
- Google sponsored “The Gapminder website http://www.gapminder.org/.

Further readings

- These documents are helpful to improve your knowledge on statistics, terminology, methods and also of limits and abuses of statistics.

o “How to Lie with Statistics” by Darrell Huff, Irving Geis (1954). W.W. Norton & Company Inc., New York.

This is a book to arm the innocent against the wiles of the statistical world, which manages to supply the reader with caution aplenty, with understandable text and amusing illustration

- “A statistical Glossary” by Valeri J. Easton and John H. McColl: http://www.stats.gla.ac.uk/steps/glossary/alphabet.html

- “Business Statistics in Practice”. Bowermann, O´Connell (2013), McGraw–Hill, 7th Edition ISBN–13: 978–0073521497

  The book places great emphasis on its use of case studies, many of which are authentic problems encountered by the authors, their colleagues or former students.

- A good example, how to convince the audience using statistics and GAPMINDER. Look at the videos by Hans Rosling:
  http://www.ted.com/talks/hans_rosling_shows_the_best_stats_you_ve_ever_seen.html

- Dennis S. Mileti (1999): Disasters by Design: A Reassessment of Natural Hazards in the United States (Natural Hazards and Disasters: Reducing Loss and Building Sustainability in a Hazardous World: A Series) – JOSEPH HENRY PRESS • 2101 Constitution Avenue, N.W. Washington D.C. 20418 ISBN 0–309–06360–4

  Disasters by Design provides an alternative and sustainable way to view, study, and manage hazards in the United States that would result in disaster–resilient communities, higher environmental quality, inter – and intra – generational equity, economic sustainability, and improved quality of life. This volume provides an overview of what is known about natural hazards, disasters, recovery, and mitigation, how research findings have been translated into policies and programs; and a sustainable hazard mitigation research agenda. Also provided is an examination of past disaster losses and hazards management over the past 20 years, including factors – demographic, climate, social – that influence loss. This volume summarizes and sets the stage for the more detailed books in the series.

6

BIBLIOGRAPHY MODULE C

INDIA, Gujarat Earthquake, Damage Assessment and Emergency, Recovery Program.

Disaster mitigation and management for West Bengal, India – An appraisal.

International Strategy for Disaster Reduction – Hyogo Framework for Action 2005 – 2015: Building the Resilience of Nations and Communities to Disasters.

INDIA: Natural Hazard Risks OCHA Regional Office for Asia Pacific Issued: 23 February 2007.

MONOGRAPH ON FLOOD MANAGEMENT – Department of Disaster Management Government of West Bengal.

Wolfers J. and E. Zitzewitz (2005): Prediction Markets Journal of Economic Perspectives –http://bpp.wharton.upenn.edu/jwolfers/Papers/Predictionmarkets.pdf

Poor people are the real losers –other–related catastrophes: Greater frequency and severity. http://www.preventionweb.net/files/2915_LosterPoorpeoplearetherealloserswelt sichtendossier52008.pdf.

Module D

Understand "Causes and Effects" of disasters described by statistics

Contents

Module D

1	**Learning Unit D–1: Introduction to the subject and test of hypothesis**	167
	1.1 What is this Module about?	167
	1.2 More on basic statistical knowledge	169
	1.2.1 Random samples of a population	169
	1.2.2 Comparison of mean	170
	1.3 Guided Exercise	172
	1.4 Independent Exercise	174
	1.5 Conclusion from the exercise	175
2	**Learning Unit D–2: Correlation – An introduction and relation to Disaster Risk Management**	177
	2.1 What is this Unit about?	177
	2.2 Correlation	178
	2.3 The scatter plot and correlation coefficient	179
	2.4 Guided Exercise	181
	2.5 Independent Exercise	183
3	**Self–Assessment for Module D**	185
4	**Useful resources**	193
5	**Bibliography Module D**	195
	Glossary	199
	List of acronyms	207

1

LEARNING UNIT D–1: INTRODUCTION TO THE SUBJECT AND TEST OF HYPOTHESIS

1.1 What is this Module about?

In this Module D we will look at possible causalities between natural or man–made causes and past disasters and future disasters under statistical aspects. Especially we will deal with forecast and predictions. We started with basic statistical analysis in Module B linking rainfall figures to floods and agricultural drought. In Module C we analysed the frequency of certain disasters and what this might mean for the capacity of knowing about future events. We used trend lines for past time lines and (implicitly) used this feature for deducing some kind of "trend" pursuing in the future. This was all without much theoretical background which will follow in this module.

However right from the start we have to highlight: in the context of forecasting and prediction. We will mostly use formal statistical methods employing time series, cross–sectional or longitudinal data.

We start with simple methods like testing of hypothesis and trend lines in order to look at the most elementary questions of forecasting. And we will not hide the reaction of many statisticians when confronted with the possibility that their profession might contribute to a discussion of causation. Many deny immediately that there is any such a possibility.

However one needs to recall that a well–designed randomized experiment can be a powerful aid in investigating causal relations. Randomized experiments have transformed many branches of science and would be able to contribute significantly to Disaster Risk Management.

You probably heard of the notion that a wing movement of a butterfly on the Amazon can "cause "a tropical storm in India. So we might better ask, what can be an "efficient cause" for a disaster?

Another example: There is no relationship between the occurrence of an earthquake and the time of the day. There has been quite some research, but it could not be proven that earthquakes at day or night time would be more frequent than at some other point in time. Also there has been research if earthquakes occur at certain constellations of the moon more frequently. And yet there is one relationship: It is between the time of the day when the earthquakes occur and the number of victims. Earthquakes occurring at night lead to higher numbers of victims, since most people are at home and sleeping when they get surprised by an earthquake and do not have the time and attention to react immediately.

Some denominations important to remember

Cause / Causality is the relationship between an event (the cause) and a second event (the effect), where the second event is a consequence of the first.

Inference is the process of drawing a conclusion by applying clues (of logic, statistics etc.) to observations or testing hypotheses; or by interpolating the next logical step in an intuited pattern. The conclusion drawn is also called an inference.

Associative inference deals with associations, relationships, correlations, etc. but not with causal connections—between variables. Causality cannot be measured, determined, or established by observation alone. Additional inputs are required to infer a causal connection or inference. So an associative inference might be causal or not?

Source: en.wikipedia.org

Expected results: the participants shall understand where fact based information tells us about relations between some observable data (e.g. rainfall) and disasters in the recent past 20 to 30 years. We will embark on mentioned various observational methods to relate to disasters. Participants will learn to distinguish between this and not inadvertently mistake this for causes. The word "Cause and Effect" are deliberately used in quotes. However participants will learn to interpret fact based information of various natural phenomena for the use of consequences in Disaster Risk Management.

Explicit actions for a Disaster Risk Management should be identified by participants while attempting the assignment.

1.2 More on basic statistical knowledge

In the previous modules we dealt with simple descriptive statistics, always keeping DRM in mind. In this part we will go some steps further trying to say something more about statistical indicators.

If we try to analyse data more thoroughly we have to introduce more statistical methods. – but we will try to keep explanations as simple as possible. Specialized textbooks may help further for in–depth knowledge.

1.2.1 Random samples of a population

Populations

A population, i.e. the quantity of all cases, about which a statement is required, is usually far too large, in order to determine for each individual case of the population the corresponding empirical data (for instance this would mean to conduct a survey asking all Indian citizens individually). In some other cases the population is simply indefinite (for instance this applies to all weather data). Another example, the size of population could be finite or infinite. The population of livestock in a district of Bihar affected by disaster may form a **Finite Population** since we may be interested to know about these livestock. But when we want to know about the fish in a river, we can term them as an **Infinite Population** of fish.

Samples

Usually one examines an entire population by drawing a sample from the original population. The fundamental assumption being that the sample proportionally reflects the same characteristics as they are present in the entire population.

This is a very important assumption.

In some cases these assumptions, though true in theory, cannot always be put into practice due to unintended errors in the way the sample is drawn. Therefore it is essential to know how to obtain undistorted statements about the population. A sample must fulfil certain requirements:

1. First of all it must be guaranteed that the sample is actually a subset of the population. Each case of the sample must be also a case of the population.
2. A second substantial condition is that it is a random sample. That means that the concrete cases of the sample are selected purely coincidentally from the population. Each case of the

population must have an equivalent chance to be included in the sample. This is neither a simple exercise to carry out, nor is it always possible to determine with certitude whether the chance of an individual of being included in the sample is really coincidental. If the information is collected, through a questionnaire then only such persons, who are ready to fill out such a questionnaire or who can read and write, will return the filled out questionnaire. Since in this approach a part of the population characterized by a specific characteristic is excluded systematically such a sample cannot guarantee distortion–free results. So, if we have a distorted sample present, then it is not possible to draw undistorted conclusions on the population, i.e. the analysis might become quite useless. Therefore it is always essential to pay particular attention to the way a sample is drawn.

This is not only the case when you conduct your own surveys, but also when you analyse the surveys of others. Next time someone presents you a survey or an analysis, we encourage you to question this person about the way the sample was drawn–up, and thereby you will be able to make up your mind about the possible quality of inference drawn out of such analysis.

1.2.2 Comparison of mean

First of all, we will understand about hypothesis and testing. Later, we will find out more about how to measure causality.

t–test

With the help of so called t–test we can assess statement about average (or mean) values of a sample. The t–test also examines whether two average values differ significantly from each other. The average values can come from individual groups of the same variables (for example the average age of men and of women) or two in pairs different variables to be taken (for example the average incomes and expenditures the same person's group). If size of sample is bigger than 30 units, one uses normal test–statistic (z–statistic) otherwise t–statistic is recommended by the statisticians.

For testing of any hypothesis, we first define the null hypothesis – this is the hypothesis we want to test against alternative hypothesis. This is done by calculating appropriate test statistic; here it is t–statistic. We name null hypothesis by the symbol Ho and an alternative hypothesis by H1. Thus in order to test the hypothesis that between two averages values no difference exists; we shall formulate the problem as Ho (the two averages are equal) and H1 (the two averages are not equal). The t–test examines, to what extent an observed difference between two average values of two samples is due to random chance or due to an “influence”.

The results of tests are accepted or rejected. If the t–test is rejected it can be concluded that also the appropriate average values from the two samples of the population are different from each other.

Let's look at two time series (time series is a sequence of data points, measured typically at successive times spaced at uniform time intervals):

> In statistics, a data point or observation is a set of one or more measurements on a single member of a statistical population. For example, in a study of the determinants of money demand with the unit of observation being the household, a data point might be the values of income, wealth, age of household head, and household size of a single household. Statistical inference about the population would be conducted using a statistical sample consisting of various such data points.

Load **RainfallIndiaLongTerm.xls,** it contains long time observations of rainfall data by months and year for many Indian Regions:

We want an answer to two simple questions.

1. Is the annual rainfall over the years different in the north east of India from the southern region?
2. Is there a difference in variation or spread of the data (variance) between these two regions?

We will probably guess the answer to the first question, yes there probably is much more rainfall in the north east than in the south. But how can we prove this, using statistics? And what about the variation, is it higher in any of the regions than in the other?

(Source: Indian Institute of Tropical Meteorology;
http://www.tropmet.res.in/static_page.php?page_id=52)

1.3 Guided Exercise

1. Look at the Tabs of "North East RF" (NORTH EAST RAINFALL SERIES FOR THE PERIOD 1829–2006.) and "Southern Peninsular RF" (SOUTH PENINSULAR RAINFALL SERIES FOR THE PERIOD 1813–2006). The Rainfall Figures are in mm.

2. We want to compare the annual figures (column ANN) for both regions using the t–test method of comparing means of two independent samples with unknown variance.

3. In our guided exercise we will refer to StatistiXL again because of more clarity in output, but the t–test is also possible in EXCEL (http://www.stattutorials.com/EXCEL/EXCEL_TTEST1.html) (last access 1.9.2013). For this in EXCEL call Tools/ Data Analysis / F–Test Two Sample for Variance. Notice there is a Tab in the File "ResT–Test StatistiXL" for storing the Results of analysis.

4. Select the annual Ranges in Figure D1.1, options are in Figure D1.2.

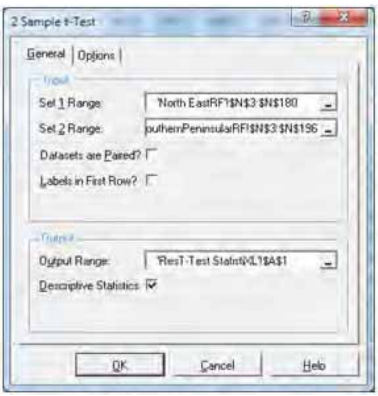


Figure D1.1

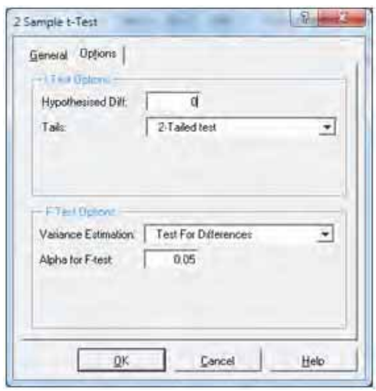


Figure D1.2

5. Press <Ok> and look at results

The header indicates that these are the results of a paired t–test, and the location of the two sets of paired data used in the analysis.

t–Test Result for Datasets:

Set 1 Range = North EastRF!$N$3:$N$180

Set 2 Range = SouthernPeninsularRF!$N$3:$N$196

Next is the descriptive statistics, giving the name of each variable, the mean, standard deviation, standard error, lower and upper confidence limit, and sample size (Notice that time range of the two series is different).

Descriptive Statistics						
Variable	**Mean**	**Std Dev.**	**Std Err**	**Lower 95% CL**	**Upper 95% CL**	**N**
Column N	2154.669	176.422	13.223	2128.573	2180.765	178
Column N	1518.292	181.020	12.996	1492.659	1543.926	194

For the 2–tailed t–test, the hypothesised difference is 0 (one of our input options) and the calculated difference in the means is 636.377 (our convention is to subtract the mean of the second group from the mean of the first, so in this example the difference in means is positive, but the sign of the difference is ignored). The standard error of the difference in the means, which is calculated from the pooled variance, is 18.562, so the t–value is 34.285 = (636.377 / 18.562) with 370 degrees of freedom (n1+n2–2). [If there had been a significant difference between the group variances, then the estimated degrees of freedom would have been calculated; every testing of hypothesis is linked with a level of significance in order to accept or reject a null hypothesis]. Finally, the probability value for this t–value given is being 0 (a more exact value is displayed in the dialog bar by clicking on the cell); we conclude that there is a significant difference between the two means of rainfall since $P < 0.05$. – The answer to our question: statistics says, that there is a difference in the rainfall mean for the two regions.

2–tailed t–Test					
Ho. Diff	**Mean Diff.**	**SE Diff.**	**T**	**DF**	**P**
0.000	636.377	18.562	34.285	370.000	0.000

Next is the F–test for equality of variance; the variance for each group is shown, and the F–ratio is calculated as the larger variance (32768.247) divided by the smaller variance (31124.833) so that $F > 1$, and its degrees of freedom are (n1–1) and (n2–1) where n1 is sample size for the group with the highest variance and n2 is the sample size for the other group. The P value for the F–ratio of 0.364 indicated that the two sample variances are not significantly different, so a pooled variance (31982.073) is used in the t–test in the example given above. The answer to our question: statistics

F–Test for Equality of Variances					
Variable	**Variance**	**F**	**DF 1**	**DF 2**	**P**
Column N	31124.833	1.053	193	177	0.364
Column N	32768.247				
Sample variances don't differ at the specified alpha of 0.0500					
so the following pooled variance was used in the t–Test.					
Pooled Variance=		31982.073			

1.4 Independent exercise

Compare from same source. Load **RainfallIndiaLongTerm.xls** and do the following exercises:

Compare means and variances of the following regions:

1. North East and North Central

Means differ and variance (spread) also differs (Right)

Means do not differ and variance (spread) also does not differ (False)

2. North East and All India

Means differ and variance (spread) also differs (Right)

Means do not differ and variance (spread) also does not differ (False)

3. North Central and North West

Means differ and variance (spread) also differs (False)

Means differ and variance (spread) does not differ (Right)

<u>Further exercise:</u>

You could also compare the means and differences of Monsoon months (JJAS = June, July, August, September) or pre–Monsoon months (MAM = March, April, May) to get more information about the flood patterns.

1.5 Conclusion from the exercise

Let's look at the test and the results:

The t–test is done to check if there is a difference of the means of the two samples, meaning that the mean of rainfall differs between the two regions so it can be assumed that the rainfall differs in general.

The 0–hypothesis is that the mean rainfalls of two regions do not differ significantly against the alternative hypothesis that they do differ significantly. If the t–test shows that the probability is close to 0 this hypothesis can be accepted (This is the case in all comparisons). If the value is less than 0.05 we will say that we accept it with a probability of 95%, if the value is less than 0.01 we will say that we accept it with a probability of 99%, the second value expressing a higher probability to accept the thesis).

If it comes to variance, the F–Test has to show results equally close to 0 to accept the hypothesis that the variance (see module 2) is the same. If we end up with a value of 0.364 like in our guided exercise it means that we will reject this hypothesis.

So our statements at the end of the exercises with the purpose to get familiarity with statistical tests and the concept of hypothesis, is for the guided exercise:

Whereas it is very likely that rainfall patterns show much different (higher) rainfall in the north east compared against the southern region the spread or variation of the rainfall seems not to have differed so much. This will be a clue for further investigation about causes of disasters like floods and cyclones.

For the independent exercises you will have to find the answers by investigating and requesting your tutor's assistance.

The formula used:

Unequal sample sizes, unequal variance.

This test is used when the two population variances are assumed to be different (the two sample sizes may or may not be equal) and hence must be estimated separately. The t statistic to test whether the population means are different can be calculated as follows:

$$t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$$\overline{X} = means$$
$$s_1^2 = Variance$$
$$n_2 = smpl.siz$$

2

LEARNING UNIT D–2: CORRELATION – INTRODUCTION AND RELATION TO DISASTER RISK MANAGEMENT

2.1 What is this Unit about?

In this unit we will explain the elementary statistical term of correlation and we will apply the related methods to analyse correlation between natural and not so natural phenomena and to disasters in India.

Correlation and dependence are any of a broad class of statistical relationships between two or more random variables or observed data values.

Familiar examples of dependent phenomena include the correlation between the physical statures of parents and their children, and the correlation between the demand for a product and its price. Correlations are useful because they can indicate a predictive relationship that can be exploited in practice. For example, an electrical utility may produce less power on a mild day based on the correlation between electricity demand and weather. Correlations can also suggest possible causal or mechanistic relationships; however statistical dependence is not sufficient to demonstrate the presence of such a relationship in case of spurious correlations. In other terms, we may use a correlation to deduce a prediction on the future behaviour. However, this is risky since the correlation does not reveal the cause or the mechanism of a relationship. The reverse will always be true: zero correlation means no dependence.

Formally, dependence refers to any situation in which random variables do not satisfy a mathematical condition of probabilistic independence. In general statistical usage, correlation can refer to any departure of two or more random variables from independence. And it is measured by coefficient of correlation or simply correlation coefficient (after Karl Pearson) sometimes also called Pearson's coefficient.

This already states one fundamental (and often misunderstood) statement of dependence: correlation is not dependence; on the contrary correlated data are excluded from causal relationships and models. If this is difficult to understand, you will learn more about it in the reasoning following the box:“ The formula used: Pearson's coefficient”.

Expected Results:

The participants will understand what the statistical terms mean and what this means for a relationship between two phenomena, in our case the two phenomena will be flood and rainfall. The first part of the “Guided Exercises” shows how to apply the correlation function to rainfall and the flood disaster figures in Bihar. In future it is assumed that participants understand to apply this procedure for any other state or district chosen in India (where data are available). It is expected that they will further understand how we chose and why we chose this scenario in Bihar and why the use of fact based information for the benefit of Disaster Risk Management is recommendable.

2.2 Correlation

Correlation is used frequently in the very general sense that two or more variable depend from each other somehow.

Let’s assume we have an indicator of the magnitude of disasters (e.g. number of casualties) and another variable, e.g. the rainfall again. Let us try to put these two variables in a scatter plot (A scatter plot is a type of graphical presentation using Cartesian coordinates to display values for two variables for a set of data. The data are displayed as a collection of points, each having the value of one variable determining the position on the horizontal axis and the value of the other variable determining the position on the vertical axis) (Source: en.wikipedia.org).

In order to obtain a first impression of a possible connection between the variables, it is convenient to present the values in a scatter plot. In a scatter plot the two values of a case determined by two variables are plotted against each other on the two scales for each case. At this arrangement; the value combinations in the diagram, you can often discover references to a connection between the variables. Frequently it shows that the points in the diagram describe the kind of the connection more appropriate as the value of the coefficient of correlation. Also with a small coefficient of correlation it is advisable to regard the values again in a scatter plot since possibly a nonlinear connection between variables is relevant which cannot be described by the coefficient of correlation. The linear Pearson coefficient of correlation is only applicable for linear relationships.

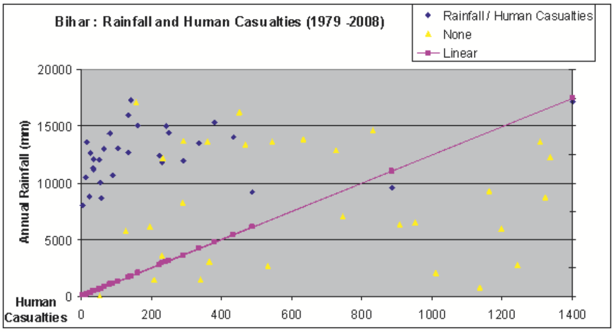


Figure D2.1: Bihar: Rainfall and human casualties (1979–2008)

Each (blue) data point represents an annual rainfall and the number of human casualties of floods in Bihar in that year (1979 – 2008), the other data points of different colour will be explained below – Example: K. Röder

Source: http://disastermgmt.bih.nic.in/Statitics/Statistics.htm and ftp://www.tropmet.res.in/pub/data/rain/iitm–subdivrf.txt

2.3 The scatter plot and correlation coefficient

In the diagram (Figure D2.1) we can observe a more or less strongly pronounced connection between the rainfall and the human victims of the flood disasters. The relationship is: The higher the rainfall, the higher the number of casualties. That is not contrary to common sense; however the connection between the rainfall and the human casualties is not particularly clear. With a very strict positive connection the points would have to be exactly on a diagonal running from left bottom to right top. Such connection is outlined in **red dots and line**. Empirically one will observe such a connection very rarely, with a strong connection the points would have to strew however at least very much closer to the diagonal (like the red dots). If no connection between the variables would exist, the points would have a scatter plot without any connection perfectly distributed over the entire diagram **(yellow dots)**.

In order to express numerically the descriptive, but not so precise impression from a scatter plot about a possible linear connection between the variables, one computes the coefficient of correlation. The Pearson coefficient of correlation is the most common.

It is widely used in the sciences as a measure of the strength of linear dependence between two variables. The correlation coefficient is sometimes called "Pearson's r".

Calculation was done with the analysis function of EXCEL: Tools/Data Analysis/Correlation

	Human casualties	Annual rainfall (mm)
Human casualties	1	
Annual rainfall (mm)	0.296324926	1

The Pearson's correlation for these two variables is 0.2963 (rounded). Compare the **Pearson's r** for the non–correlated series (yellow) is 0.038 and of course it is 1 for the red series.

An assumption of correlation is that the both of the variables (the measurements) be of continuous data measured on an interval/ratio scale. Also, each variable should be approximately normally distributed. Both preconditions are met, being collected data responding to the central limit theorem.

If the correlation is positive, it means that when one variable increases, the other tends to increase. If the correlation is negative, it means that when one variable increases, the other tends to decrease. When a correlation coefficient is close to +1 (or –1), it means that there is a strong correlation – the points are scattered along a straight line. For example, a correlation r = 0.7 may be considered strong. However, the closer a correlation coefficient gets to 0, the weaker the relationship, where the cloud (scatter) of points is not close to a straight line.

So we conclude from plot and calculation, that there is a weak relationship between the two variables but far from the assumption that rainfall alone is "causing" floods in Bihar. (A correlation greater than 0.8 would be described as strong, whereas a correlation less than 0.5 would be described as weak

Source: http://www.bized.co.uk/timeweb/crunching/crunch_relate_expl.htm).

We will continue with a guided exercise on rainfall data on state level of India and related natural disaster.

2.4 Guided Exercise

1. Go to the appropriate data page for rainfall data and results of disasters, e.g. Flood disasters: Flood casualties in Bihar: http://en.wikipedia.org/wiki/Floods_in_Bihar#Statistics and http://disastermgmt.bih.nic.in/.

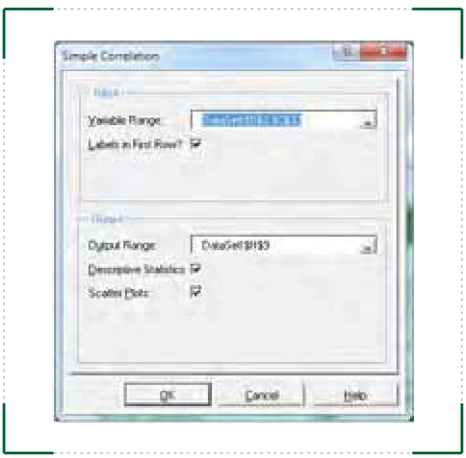
   Rainfall data: ftp://www.tropmet.res.in/pub/data/rain/iitm–subdivrf.txt

2. Open the rainfall data and transfer Bihar data (data for Indian regions are identified by Name (e.g. look for Bihar) copy all lines containing Bihar data to EXCEL, transfer text to EXCEL cells. If you don't know how to do it refer to a textbook or an online tutorial (e.g. http://www.datapigtechnologies.com/flashfiles/texttocolumn.html or more general EXCEL tutorials in http://datapigtechnologies.com/ExcelMain.htm).

3. The data file Rainfall&LossesBihar.xls has been prepared to be used for the guided exercise.

4. Call StatistiXL/Correlation/Simple.

The following box will appear and has to be filled out accordingly.



5. We click <Ok> and the results show in the indicated cell

Pearson Correlation Results for:

Variable Range = DataSet!$B$3:$C$33

First, a header is provided, which indicates the type of test: **"Pearson Correlation Results for"** and the data range.

Descriptive Statistics				
Variable	Mean	Std Dev.	Std Err	N
Human Casualties	214.800	292.444	53.393	30

The descriptive statistics table gives the mean, standard deviation, standard error and sample sizes for each variable.

Correlation Matrix ®				
	Human Casualties	**Annual Rainfall (mm)**		
Human Casualties	1.000	0.296		
Annual Rainfall (mm)	0.296	1.000		
Annual Rainfall (mm)	12509.100	2453.207	447.892	30

The correlation matrix gives the correlation coefficient (r) for all of the combinations of the variables; r is 1.000 for any variable correlated with itself, and is >= –1 and <= +1 for all other combinations e.g. r is 0.296 for the correlation of rainfall and human casualties. Note that the r value for the correlation of human casualties and rainfall is exactly the same (0.296).The corresponding correlation values (variable a to b and b to a) are always identical. This is due to the simple matrix operation and so this is true for all the other subsequent matrix values shown below.

Statistic		
	Human Casualties	**Annual Rainfall (mm)**
Human Casualties	–	1.642
Annual Rainfall (mm)	1.642	–

A simple correlation value may have a different meaning for different data sets. Imagine a data series with 4 observed values against another one with 40000, both showing the same correlation coefficient. The statistical probability of the coefficient (the t–statistic) relates this value to the distribution of the variable and the probability is calculated that this value is purely arbitrary. The matrix of t values associated with the correlation coefficients is used to determine the statistical probability of each correlation e.g. the t value for the correlation coefficient of rainfall and human casualties is 1.642 and this is used to determine the P (probability) value for the same relationship presented in the next table.

Correlation Significance (P)		
Human Casualties	**Annual Rainfall (mm)**	
Human Casualties	–	0.112
Annual Rainfall (mm)	0.112	–

The matrix of significance levels for the correlation coefficients gives the probability of obtaining the observed r value for correlation e.g. the significance level for the correlation coefficient of Annual Rainfall and Human Casualties is 0.112. Since the correlation is poor we are not sure if we can rely on this with high probability (our Null–hypothesis Ho=0 would be accepted only with 89% probability).

2.5 Independent Exercise

Compare from same source. Load HYPERLINK "Exercises/RainfallBihalITM.xls" and do the following exercises, You will have to find the answers to the following questions, answering them with **Right** or **False**

Compare correlation coefficients of flood disaster results and the following rainfalls:

1. Human casualties and Monsoon rainfall

a) The correlation of monsoon rainfall and human casualties is higher than the correlation of annual rainfall and human casualties

b) The correlation of monsoon rainfall and human casualties is lower than the correlation of annual rainfall and human casualties

2. Animal losses and Annual and Rainfall a) The correlation of animal losses and monsoon rainfall is lower than the correlation of annual rainfall and human casualties b) The correlation of animal losses and monsoon rainfall is higher than the correlation of annual rainfall and human casualties c) The correlation of animal losses and rainfalls (monsoon or annual) is higher than the correlation of rainfall (monsoon or annual) and human casualties
Answers 1. a) 2. a)

3

SELF–ASSESSMENT MODULE D

Question 1

Statistic terminology

What is statistical inference?

Select the correct answers.

a) Drawing conclusions by applying clues of logic or statistics to observations or hypotheses

b) Free association

c) For lacking verbal argumentation, we choose statistical formulas

d) Associative inference deals with associations, relationships, correlations, etc.—but not the casual connections—between variables

Question 2

More on Statistical terminology

Which of the following examples are randomized samples?

Select the correct answers.

a) Rainfall data delivered by the Meteorological Institute

b) Traffic Reports by Road Police

c) Reports on earthquake Damages

d) HIV prevalence measured by reports of hospital cases

Question 3:

About statistical tests

Which of the following statements are true?

Select the correct answers.

a) A t–test can make inference about the difference of two means.

b) A t–test is any statistical hypothesis test in which the test statistics follow a Students distribution.

c) The Student`s t distribution is independent of the sample size.

d) T–Test and F–Statistics are the same.

Question 4:

Relationship among variables

What does the correlation measure? Which of the following statements are true?

Select the correct answers.

a) The correlation varies between –1 and +1.

b) The correlation is the observed relationship among the values of a first variable and the values of a second variable.

c) If a correlation coefficient has a value close to –1. This means that there is no observed relationship among the variables.

d) Correlation and Covariance is the same.

Question 5:

About Regression

What does the statistical regression measure? Which of the following statements are true?

Select the correct answers.

a) Regression explains the cause of one event by observing others.

b) The linear regression explains the variation in one variable, a dependent numerical variable in terms of one or more predictor variables.

c) If the regression equation expresses mathematically a linear relationship, we speak of linear regression.

d) With regression functions we can predict the future.

e) Regression functions allow predictions under certain statistical preconditions.

Question 6:

About Confidence intervals

Which of the Following remarks are correct?

Select the correct answers.

a) Confidence intervals relate the quality of predictions with observed distribution of the sample.

b) It is better not to publish confidence intervals because nobody understands them anyway.

c) Damages by floods in urban and rural areas are very similar.

d) The Further we predict in the future, the wider the confidence interval becomes, the more unreliable the predictions will be.

e) Confidence intervals are always related to a confidence level (a probability).

f) The 99% confidence interval is `narrower` than the 95% interval for the same sample.

Question 7:

The results of a regression analysis

Which of the following remarks are correct?

Select the correct answers.

a) The coefficient of determination is an indicator of the `quality` of the relationship of dependent and independent variables.

b) The regression is analysed by a statistician according to his/her experience and rule of a thumb.

c) The more independent (predictor) variables are selected; the better will be the prediction based on the regression analysis.

d) The selection of explaining variables requires skill, sincerity and common reasoning.

e) Since regression is so complicated, use and interpretation should be left to the expert.

Question 8:

Interpretation of the equation of linear regression

If we want to interpret the equation of linear regression: which of the following remarks are correct?

Select the correct answers.

a) We say, that if the dependent variables have small correlation, the dependent variable can be predicted.

b) If all regression coefficients are set to 0 the intercept is the value of the dependent variable.

c) Partial Regression Coefficient indicates how much the dependent variables will change per unit change in the corresponding independent variable if all other independent variables are held constant.

d) The more sources you compare the better the chances are for better Disaster risk Management.

e) A good result of the regression means adjusted R2 close to 1.

Question 9:

About Interpretation of Statistical results

If we want to interpret statistical results, which of the following remarks are correct?

Select the correct answers.

a) We say that once an official index is published, it must be correct, even if we don't know much about the calculation of this index.

b) A majority difficulty about statistical results for Disaster Risk Management is to define and measure the `precise outcome` of a disaster because the outcomes of a disaster are often multiple.

c) The interpretation of statistical results should improve the reasoning for more profound research to improve Disaster Risk Management in general.

d) The current use of the interpretation of statistical results is sufficient for Disaster Risk Management in general.

e) Insight into the calculation of an earthquake risk index will increase awareness of the different components of what contributes to this risk.

Question 10:

On factor and Principal Component Analysis

If we want to distinguish Factor and Principal Component Analysis, which of the following remarks are correct?

Select the correct answers.

a) Principal component analysis (PCA) involves a mathematical procedure that transforms a number of possibly correlated variables into a smaller number of uncorrelated variables called principal components.

b) In factor analysis, we look for common factors based on external assumptions.

c) Both methods should be mixed and simultaneously used.

d) Both methods can be mixed as long it is explained which method and which mathematical approach is used.

e) In factor analysis there is a structured model which resembles the linear regression approach.

Question 11:

The reasoning behind Earthquake Risk Index

What is the reasoning behind an Earthquake Risk Index, which of the following remarks are correct?

Select the correct answers.

a) The effects or damages occurred after an earthquake depends on many factors like infrastructure and compositions of the population.

b) An earthquake risk Index should rely entirely on the past seismicity (occurrence of earthquakes) of an area.

c) In the reasoning for an earthquake Risk Index, the different factors ` Geological conditions` and Tsunami potentials` were omitted.

d) The information about various components of earthquake risk is readily available for the Indian cities.

e) The higher the percentage of wooden constructions, the higher the potential hazard by post–earthquakes fires.

Question 12:

Further thoughts about an Indian urban Earthquake Risk index

What if there are further thoughts about an Indian urban Earthquake Index, Which of the flowing remarks are correct?

Select the correct answers.

a) The data quality, the model and the calculation method should be refined with robust and agreed data series.

b) The Earthquake risk Index should even be simpler, the vulnerability of the population and the past seismicity (occurrence of earthquakes) of an urban area will be sufficient.

c) In the reasoning for an earthquake Risk index, The different factors ` Geological conditions` and ` Tsunami potentials` were omitted.

d) Disaster risk management should react to specific weakness indicated by partial indices.

Answers

1. a & d
2. a & b
3. a
4. a & b
5. b, c & e
6. a, d & e
7. a & d
8. b, c & e
9. b, c & e
10. d & e
11. a & e
12. d

USEFUL RESOURCES

The following online and offline resources will need to be accessed in order to complete the task. The orientation is mainly done through the previous modules showing examples and best practices how data can improve knowledge and handling of natural and industrial risks for the benefit of Disaster Risk Management. Which resources (data) needed will mainly be decided by the case study. However, there are some common resources, like the EXCEL template, which are mandatory to use and a participant can consider others like references to general statistical information sources.

Mandatory tools, media and learning material to use

- EXCEL template to be filled with existing, relevant, missing data essential for the task. A template for this is prepared.

  (Mod4_Template_EXCEL).

Optional tools and media

- Download a full working version of StatistiXL which includes all Statistical modules, Help files and Example files and will run for 30 days, allowing you to check out all of its features with your own data. After this time, you will need to purchase a key from our online store if you want to continue using StatistiXL.

  http://www.statistixl.com

  StatistiXL greatly extends this feature set to encompass high powered statistical analysis without the need to learn how to use an entirely new application from scratch tool.

Further readings

These documents are helpful to improve your knowledge on statistics, terminology, methods and also of limits and abuses of statistics.

- Huff D., I. Geis (1954): How to Lie With Statistics. W.W. Norton & Company Inc., New York.

  Here is a book to arm the innocent against the wiles of the statistical world, which manages to supply the reader with caution aplenty, with understandable text and amusing illustration.

- Connell, Bowermann, O´(2003): Business Statistics in Practice, McGraw–Hill, 3rd Edition.

  The book places great emphasis on its use of case studies, many of which are authentic problems encountered by the authors, their colleagues or former students.

- Mileti D.S. (1999): Disasters by design: a reassessment of natural hazards in the United States. Joseph Henry Press 2101, Washington D.C.

- Easton V. J., & McColl, J. H. A statistical Glossary.
  http://www.stats.gla.ac.uk/steps/glossary/alphabet.html.

- Hans Rosling: A good example, how to convince the audience using statistics and gap minder
  http://www.ted.com/talks/hans_rosling_shows_the_best_stats_you_ve_ever_seen.html.

3

BIBLIOGRAPHY MODULE D

Sources used for the text

Davidson, Shah (1997): An Urban Earthquake Disaster Risk Index – Stanford http://blume.stanford.edu/pdffiles/Tech%20Reports/TR121_Davidson.pdf.

Filmer D., L. Pritchett (1998): Estimating Wealth Effects without Expenditure Data Download: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=620538.

Carvalho Soniya, W. Howard (1996): Combining Qualitative and Quantitative approaches to Poverty Measurements and Analysis. World Bank Technical Paper No. 366, Washington DC.

Ravallion M. (1992): Poverty Comparisons: A guide to concepts and Methods. The World Bank.

Ravallion M. (1996): Issues in measuring and Modelling Poverty. The Economic Journal, Vol. 106, No 438 (Sep., 1996), pp 1328–1343.

Ravallion M. (2003): The Debate on Globalization, Poverty, and Inequality: Why Measurement Matters. Policy Research Working Paper 3038; The World Bank.

Ravallion M. (2003): Have We Already Met the Millennium Development Goal for Poverty? http://iiea.iie.com/publications/papers/ravallion0203.pdf.

WB–Using Census and HH Data for Poverty. http://web.worldbank.org/WBSITE/EXTERNAL/TOPICS/EXTPOVERTY/EXTPA/0,,contentMDK:21094736~isCURL:Y~menuPK:435761~pagePK:148956~piPK:216618~theSitePK:430367,00.html.

IFPRI–Using Census and HH Data for Poverty. ..\Docs\IFPR–Census use– Poverty, Inequality, and Geographic Targetingfcndp192.pdf.

Costs of Household Surveys in SS–Africa.
http://unstats.un.org/unsd/hhsurveys/pdf/Chapter_14.pdf.

Measuring Human Development – HDI.
http://hdr.undp.org/en/reports/publications/title,4182,en.html.

Klaus Röder–Project Proposal for a Survey Measuring Development in Sofala on District Level. ..\Docs\ProposingSMD_Sofala070301_...pdf.

Simler K.R., S. Mukherjee, G.L. Dava and G. Datt (2003): IFPRI–Research Report 132 Rebuilding after War Micro–level Determinants of Poverty Reduction in Mozambique IFPRI 20006–1002 USA. http://www.ifpri.org/pubs/abstract/132/rr132.pdf.

Estimating Wealth Effects without Expenditure Data.
http://www.worldbank.org/html/dec/Publications/Workpapers/WPS1900series/wps1994/wps1994.pdf.

Klaus Röder (2007): Comparing the Results over 3 Years: O DesenvolvimentoDistritalemSofalaatravés do Inquérito SMD: Resultados 2005/2006/2007 (port.).
\../Docs/Resultados_SDM_Sofala070522_PreGTZ_PT.doc.

Holzinger K.J., H.H. Harman (1967): Factor Analysis: A Synthesis of Factorial Methods. Vol. 75, No. 1, Spring.

Status of Earthquake Vulnerability and Disaster Risk Management in Himachal Pradesh. Dr. G.P. Kapoor, Professor and Head (DMC) Shimla, Himachal Pradesh; 2011.

Statistical handbook for the course Disaster Risk Management and Statistics, compiled for GIZ by Klaus Röder – 2011 – DRM Statistics Handbook_final_101117.pdf.

Further readings

These documents are helpful to improve your knowledge on statistics, terminology, methods and also of limits and abuses of statistics.

- "How to Lie with Statistics" by Darrell Huff, Irving Geis (1954). W.W. Norton & Company Inc., New York.

This is a book to arm the innocent against the wiles of the statistical world, which manages to supply the reader with caution aplenty, with understandable text and amusing illustration.

- SPSS 10.0 Regression Models ISBN 0130179043.
- Brosius G. (1989): SPSS– Advanced Statistics and Tables McGraw–Hill Book Company Inc.

GLOSSARY

Average	An average is a single value that is meant to typify a list of values. The most common method is the arithmetic mean. Suppose we have a data set containing the values The arithmetic mean is defined by the formula . $A = \frac{1}{n}\sum_{i=1}^{n} a_i$. The arithmetic mean, often simply called the mean, of two numbers, such as 2 and 8, is obtained by finding a value A such that 2 + 8 = A + A. One may find that A = (2 + 8)/2 = 5.
Data	The term **data** refers to qualitative or quantitative attributes of a variable or set of variables. Data (plural of **"datum"**) are typically the results of measurements and can be the basis of graphs, images, or observations of a set of variables. Data are often viewed as the lowest level of abstraction from which information and then knowledge are derived
Database	A **database** is an organized collection of data for one or more purposes, usually in digital form. The data are typically organized to model relevant aspects of reality (for example, the availability of rooms in hotels), in a way that supports processes requiring this information (for example, finding a hotel with vacancies). The term "database" refers both to the way its users view it, and to the logical and physical materialization of its data, content, in files, computer memory, and computer data storage. This definition is very general, and is independent of the technology used. However, not every collection of data is a database; the term database implies that the data is managed to some level of quality (measured in terms of accuracy, availability, usability, and resilience) and this in turn often implies the use of a general–purpose Database management system (DBMS). A general–purpose DBMS is typically a complex software system that meets many usage

	requirements, and the databases that it maintains are often large and complex. The utilization of databases is widely spread to a degree that virtually any technology and product nowadays rely on databases and DBMSs for their development and commercialization, or even may have such embedded in them. Also organizations and companies, from small to very large, heavily depend on databases for their operations.
Disaster	A disaster is a catastrophe, mishap, calamity in any area, arising from natural or manmade causes, which results in substantial loss of life or human suffering, damage destruction of, property, environment, and is of such a nature or magnitude as to be beyond the coping capacity of the community of the affected area. Impacts of Disasters are a) Loss of lives b) Loss to Property and infrastructure c) Damage to livelihood d) Economic Losses e) Environmental Damage– Flora & Fauna f) Sociological & Psychological after effects g) Civil Strife
Eigenvalue	In mathematics, Eigenvalue, Eigenvector, and Eigenspace are related concepts in the field of linear algebra. Eigenvalues, Eigenvectors and Eigenspaces are properties of a matrix. They are computed by a method described below, give important information about the matrix, and can be used in matrix factorization. They have applications in areas of applied mathematics as diverse as economics and quantum mechanics.

	In general, a matrix acts on a vector by changing both its magnitude and its direction. However, a matrix may act on certain vectors by changing only their magnitude, and leaving their direction unchanged (or possibly reversing it). These vectors are the eigenvectors of the matrix. A matrix acts on an eigenvector by multiplying its magnitude by a factor, which is positive if its direction is unchanged and negative if its direction is reversed. This factor is the Eigenvalue associated with that eigenvector. An Eigenspace is the set of all eigenvectors that have the same Eigenvalue, together with the zero vector. These concepts cannot be formally defined without prerequisites, including an understanding of matrices, vectors, and linear transformations.
Equation	An equation is a mathematical statement that asserts the equality of two expressions. Equations consist of the expressions that have to be equal on opposite sides of an equal sign.
Hazard	Hazard is a potentially damaging physical event, phenomenon or human activity that may cause loss of life or injury, property damage, social and economic disruption or environmental degradation. Hazards are classified on the basis of the origin i.e. Natural (geological, hydro–meteorological and biological), human induced process (environmental degradation and technological hazards). Hazards can be single, sequential or combined in their origin and effects. Each hazard is characterized by its location, intensity, frequency and probability.
Hypothesis	Setting up and testing hypotheses is an essential part of statistical inference. In order to formulate such a test, usually some theory has been put forward, either because it is believed to be true or because it is to be used as a basis for argument, but has not been

	proved, for example, claiming that a new drug is better than the current drug for treatment of the same symptoms. In each problem considered, the question of interest is simplified into two competing claims / hypotheses between which we have a choice; the null hypothesis, denoted H0, against the alternative hypothesis, denoted H1. These two competing claims / hypotheses are not however treated on an equal basis: special consideration is given to the null hypothesis. We have two common situations: The experiment has been carried out in an attempt to disprove or reject a particular hypothesis, the null hypothesis, thus we give that one priority so it cannot be rejected unless the evidence against it is sufficiently strong. For example, H0: there is no difference in taste between coke and diet coke against H1: there is a difference.
Interval	In mathematics, a (real) interval is a set of real numbers with the property that any number that lies between two numbers in the set is also included in the set.
Kaiser normalisation	Kaiser normalisation is commonly used to normalise the initial factor loadings by dividing by the communality to remove bias in rotation; and the rotated factor loadings are then un–normalised.
Kurtosis	In statistics, kurtosis is a measure of the "peakedness" of the probability distribution of a real–valued variable. Higher kurtosis means more of the variance is the result of infrequent extreme deviations, as opposed to frequent modestly sized deviations.

Monte Carlo Simulation	Monte Carlo methods are a class of computational algorithms that rely on repeated random sampling to compute their results. Monte Carlo methods are often used in Computer simulations of physical and mathematical systems.
Normal distribution, Normality	The normal distribution is often used to describe, at least approximately, any variable that tends to cluster around the mean. For example, the heights of adult males in the United States are roughly normally distributed, with a mean of about 70 inches (1.8 m). Most men have a height close to the mean, though a small number of outliers have a height significantly above or below the mean. A histogram of male heights will appear similar to a bell curve, with the correspondence becoming closer if more data are used.
	By the central limit theorem, under certain conditions the sum of a number of random variables with finite means and variances approaches a normal distribution as the number of variables increases. For this reason, the normal distribution is commonly encountered in practice, and is used throughout statistics, natural sciences, and social sciences as a simple model for complex phenomena. For example, the observational error in an experiment is usually assumed to follow a normal distribution, and the propagation of uncertainty is computed using this assumption.
Official statistics	Official statistics are statistics published by government agencies or other public bodies such as international organizations. They provide quantitative or qualitative information on all major areas of citizens' lives, such as economic and social development, living conditions, health, education, and the environment.
Orthogonal	In mathematics, two vectors are orthogonal if they are perpendicular, i.e., they form a right angle.

Population	In mathematics, statistics, and the mathematical sciences, a parameter is a quantity that serves to relate functions and variables using a common variable (often t) when such a relationship would be difficult to explicate with an equation. Example: for a population of test scores, a parameter would not be an actual score, but perhaps an average computed from all scores, or a per cent computed from all scores.
Probability distribution	In statistics, a probability distribution identifies either the probability of each value of a random variable (when the variable is discrete), or the probability of the value falling within a particular interval (when the variable is continuous).The probability distribution describes the range of possible values that a random variable can attain and the probability that the value of the random variable is within any (measurable) subset of that range.
Random variable	In mathematics, a random variable (or stochastic variable) is a way of assigning a value (often, a real number) to each possible outcome of a random event. Intuitively, a random variable can be thought of as a quantity whose value is not fixed, but which can take on different values according to a probability distribution. These different values might represent, for example, different outcomes of a random experiment (throwing dice).
Risk	Risk is the (so far non–statistical) probability of harmful consequences or expected losses (deaths, injuries, property, livelihoods, economic activity disrupted or environment damaged) resulting from interaction between natural or human–induced hazards and vulnerable conditions.
Rotation	A rotation is a movement of an object in a circular motion. An object rotates around a centre (or point) of rotation.

Secondary data	**Secondary** data is data collected by someone other than the user. Common sources of secondary data for social science include censuses, surveys, organizational records and data collected through qualitative methodologies or qualitative research. Primary data, by contrast, are collected by the investigator conducting the research.
Surge (Tropical surge, coastal surge und storm surge)	**Storm surge** is an offshore rise of water associated with a low pressure weather system, typically a tropical cyclone. Storm surge is caused primarily by high winds pushing on the ocean's surface. The wind causes the water to pile up higher than the ordinary sea level. (Source: en.wikipedia.org)
Trend line	A **trend line** represents a trend, the long–term movement in time series data after other components have been accounted for. It tells whether a particular data set (say GDP, oil prices or stock prices) have increased or decreased over the period of time. A trend line could simply be drawn by eye through a set of data points, but more properly their position and slope is calculated using statistical techniques like linear regression. Trend lines typically are straight lines, although some variations use higher degree polynomials depending on the degree of curvature desired in the line.
Tsunami	A **Tsunami** or tidal wave is a series of water waves caused by the displacement of a large volume of a body of water, such as an ocean or a large lake. Earthquakes, volcanic eruptions and other underwater explosions, landslides and other disturbances above or below water all have the potential to generate a tsunami. (Source: en.wikipedia.org)

Variability	In statistics, statistical dispersion (also called statistical variability or variation) is variability or spread in a variable or a probability distribution. Common examples of measures of statistical dispersion are the variance, standard deviation and inter–quartile range. Dispersion is contrasted with location or central tendency, and together they are the most used properties of distributions. (Source: en.wikipedia.org)
Vulnerability	The conditions determined by physical, social, economic and environmental factors or processes, which increase the susceptibility of a community to the impact of hazards, are called vulnerability. Vulnerability can be due to the following factors a) Geographic and Environmental b) Physical c) Social d) Economic e) Systemic

LIST OF ACRONYMS

ADRC	Asian Disaster Reduction Centre
AFP	Agence France–Press
APELL	Awareness and Preparation for Emergencies on a Local Level
AWiFS	Advanced Wide Field Sensor
BARPI	Bureau d'Analyses des Risques et des Pollutions Industrielles of the French Ministry of the Environment
BBC	British Broadcasting Corporation
BCPR	Bureau for Crisis Prevention and Recovery
BMTPC	Building Material Technology Promotion Council
BSDMA	Bihar State Disaster Management Authorities
CDD	Canadian Disaster Database
CNN	Cable News Network
CRED	Centre for Research on the Epidemiology of Disasters (WHO)
CSO	Central Statistical Organisation
CSU	Central Surveillance Unit
CWC	Central Water Commission
DMSP	Defence Meteorological Satellite Program
DOS	Department of Space
DRM	Disaster Risk Management

DROMIC	Disaster Response Operations Monitoring & Information Centre
DRR	Disaster Risk Reduction
DSC	Decision Support Centre)
DWSD	Department of Social Welfare and Development
EMA	Disasters Database (Emergency Management Australia)
EM–DAT	Emergency Disasters Data Base
ENVIS	Environmental Information System
FAO	Food and Agriculture Organization (United Nations)
GDAC	Global Disaster Alert and Coordination System
GLIDE	Global Disaster Identification Number
GSI	Geological Survey of India
HPC	High Power Committee
ICT	Information and Communication Technology
IDRN	India Disaster Resource Network
IDSP	Integrated Disease Surveillance Project
IFRC	Advanced Very High Resolution Radiometer
IMD	India Meteorology Department
INFFRAS	Indian Forest Fire Response and Assessment System

IRIS	Incorporated Research Institutions of Seismology
IRS	Indian Remote Sensing Satellite
KSDMA	Karnataka State Disaster Management Authorities
MHIDAS	Major Hazard Incident Data Service
MoA	Ministry of Agriculture
MODIS	Moderate Resolution Imaging Spectro–radiometer
MoSPI	Ministry of Statistics and Programme Implementation
NADAMS	National Agricultural Drought Assessment and Monitoring System
NASA	National Aeronautics and Space Administration
NCDC	National Civil Defence College, Nagpur
NCDC	US National Climatic Data Centre
NCPP	Coastal Protection Project
NCT	National Capital Territory
NDCC	National Disaster Coordinating Council
NEMO	National Emergency Management Organization
NGO	Non Governmental organisation
NIC	National Informatics Centre
NIDM	National Institute of Disaster Management

NOAA	National Oceanic and Atmospheric Administration
AVHRR	Advanced Very High Resolution Radiometer
NRSC	National Remote Sensing Centre
NSSO	National Sample Survey Organisation
NWIP	National Wastelands Inventory Project
OCHA	UN Office for the Coordination of Humanitarian Affairs
OCIPEP	Office of Critical Infrastructure Protection and Emergency Preparedness
OECD	Organisation for Economic Co–operation and Development
OFDA	Office of Foreign Disaster Assistance
RTSMN	Real Time Seismic Monitoring Network
SDMA	State Disaster Management Authorities
SDR	State Human Development Reports
SHELDUS	Spatial Hazard Event and Losses Database for the United States
UNDP	United Nations Development Programme
UNEP	United Nations Environmental Program
WFP	World Food Programme
WiFS	Wide Field Sensor

About NIDM

National Centre for Disaster Management (NCDM) set up under the Department of Agriculture and Cooperation, Ministry of Agriculture in March 1995. NCDM has been upgraded into full–fledged National Institute of Disaster Management in October 2003. Under the Disaster Management Act, 2005, the Institute has been entrusted with the nodal national responsibility for human resource development, capacity building, training, research, documentation and policy advocacy in the field of disaster management.

NIDM is steadily marching forward to fulfil its mission to make a disaster resilient India by developing and promoting a culture of prevention and preparedness at all levels. Both as a national centre and then as the national Institute, NIDM has performed a crucial role in bringing disaster risk reduction to the forefront of the national agenda. It is our belief that disaster risk reduction is possible only through promotion of a "Culture of Prevention" involving all stakeholders.

We work through strategic partnerships with various ministries and departments of the central, state and local governments, academic, research and technical organizations in India and abroad and other bi–lateral and multi–lateral international agencies.

About GIZ

The services delivered by the Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH draw on a wealth of regional and technical expertise and tried and tested management know–how. As a federal enterprise, we support the German Government in achieving its objectives in the field of international cooperation for sustainable development. We are also engaged in international education work around the globe. GIZ currently operates in more than 130 countries worldwide.

GIZ in India

Germany has been cooperating with India by providing expertise through GIZ for more than 50 years. To address India's priority of sustainable and inclusive growth, GIZ's joint efforts with the partners in India currently focus on the following areas:

- Energy – Renewable Energy and Energy Efficiency
- Sustainable Urban and Industrial Development
- Natural Resource Management
- Private Sector Development
- Social Protection
- Financial Systems Development
- HIV/AIDS – Blood Safety

About the Indo–German Environment Partnership (IGEP) Programme

IGEP builds on the experience of the predecessor Advisory Services in Environment Management (ASEM) programme but at the same time strengthens its thematic profile in the urban and industrial sector, up–scales successful pilots and supports the environmental reform agenda and priority needs of India.

The overall objective of IGEP is that the decision makers at national, state and local level use innovative solutions for the improvement of urban and industrial environmental management and for the development of an environment and climate policy that targets inclusive economic growth de–coupled from resource consumption.

For information visit http://**www.igep.in** or write to **contact@igep.in.**

Authors Profile

Sreeja S. Nair

Sreeja S. Nair is Assistant Professor at National Institute of Disaster Management since 2007. She is disaster management professional having more than 12 years of experience in the field. Her areas of research, documentation and training activities at NIDM include geoinformatics applications in disaster management, environmental law, disaster data and information management, ecosystem approach to disaster risk reduction and chemical disaster management. Ms. Nair published 12 papers in national and international journals, authored 5 training modules and editor of 3 books. She is the coordinator of Indo German Cooperation on Environmental Knowledge for Disaster Risk Management and co–principal investigator of ICSSR research project on drought vulnerability and mitigation analysis. She is also involved as a technical expert in the GIZ–European Union pilot project on integrating climate–change adaptation with disaster management planning process coastal Andhra & Tamil Nadu.

Dr. Anil K. Gupta

Dr. Anil K. Gupta, Head of Policy Planning Division and Training Section, joined the National Institute of Disaster Management in 2006 as Associate Professor. His areas are disaster management, climate–change adaptation and NRM, with focus on risk/vulnerability assessment, planning and management. He did his Post–Doctoral work at NEERI Nagpur (CSIR) and received Young Scientist Award in 1996. He is Fellow of Earth Scientists Society. He was Reader & Head of Department of Environment & NRM since 2003 and founder Director of Institute of Environment & Development Studies at Bundelkhand University. Earlier he worked at the Ambedkar Central University of Lucknow, Disaster Management Institute (Govt. of MP, Bhopal), National Mineral Development Corporation, and CICON Environment Technologies. He has supervised several Ph.D. and PG research, has over 100 publications including 3 books, 4 training modules and 44 papers in refereed journals. He implemented several research/coordinated projects supported by GIZ Germany, UNDP, UNEP, CDKN, ICSSR, and MoEF, focussing on drought, floods, climatic–risks, coastal hazards, DM planning, environmental knowledge and legislation. He has 23 experiences of research including 17 years teaching/training and 8 years administrative experience.

Klaus Röder

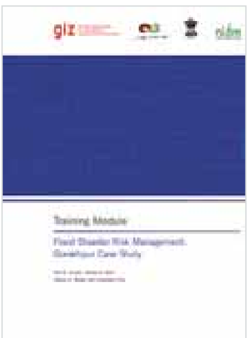
Klaus Röder has a Master's degree in Mathematics, Computers Science and Applied Statistics at Technical University in 1977. Worked in the early 1980s with the Munich Centre for post–graduate Statistical Training of the Carl Duisberg Society in Germany as Program Director and Training Expert for Computers and Statistics and as independent consultant for Statistics and Computer Applications. Project leader for the Munich Centre promoting evidence based policy making in Sub Saharan Africa in Malawi, Mali, Ruanda, Tanzania and Zimbabwe, closely cooperating within a EUROSTAT team with other European development oriented organisations like Leuven University, INSEE Paris and the Oxford Food Studies Group, project coordinator in South America especially in Paraguay and Brazil and as a business consultant in the position of head of department in Europe (now NTT Data – Munich branch). From 2003 to 2007 he worked for the Statistical Office in Mozambique.

Special field of interest is the measuring and monitoring of environmental, economic and policy decisions, like in business surveys in Paraguay and the impact of tourism on regional economic and environmental development in Brazil and lately with a tailored survey in Mozambique: measuring prosperity and well–being on district level in Mozambique in the most recent years. From 2007 until today he is an independent consultant for various development projects in several countries (Indonesia, Ghana, Nigeria, Liberia, and Syria etc.) usually related to Statistics or Monitoring results of development plans. In India he supported the guidelines for the inclusion of statistical environmental data into a capacity development system which should enhance the Disaster Risk Management in India. He has developed the Blended Learning course on Knowledge base and Statistics for Disaster Risk Management and conducted jointly with NIDM and GIZ in 2010 and a "Training of Trainers course" in 2011.

Environmental Knowledge and Disaster Risk Management
Training modules

Environmental Legislation for Disaster Risk Management
ISBN: 978-3-944152-12-7
Anil K. Gupta,
Sreeja S. Nair and S. Singh

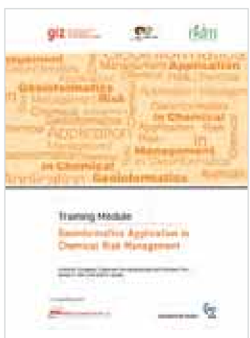
Flood Disaster Risk Management-Hydrological Forecasts: Requirements and Best Practices'
ISBN: 978-3-944152-15-8
A. Vogelbacher,
Bavarian Water Agency



Flood Disaster Risk Management: Gorakhpur Case Study
ISBN: 978-3-944152-14-1
Anil K. Gupta,
Sreeja S. Nair,
Shiraz A. Wajih and
Sunanda Dey

Databases and Statistics for Disaster Risk Management
(4 Modules)
ISBN: 978-3-944152-11-0
Sreeja S. Nair, Klaus Röder
& Anil K. Gupta

Critical infrastructures and Disaster Risk Reduction in the Light of Natural Hazards
ISBN: 978-3-944152-13-4
Claudia Bach,
Anil K. Gupta,
Sreeja S. Nair and
Jörn Birkmann



Geoinformatics Application in Chemical Risk Management
ISBN: 978-3-944152-33-2
Anandita Sengupta,
Debanjan Bandyopadhyay,
Nilanjan Paul,
Sreeja S. Nair and
Anil K. Gupta

Indo–German Environment Partnership (IGEP) programme,
GIZ India
&
National Institute of Disaster Management, Government. of India, New Delhi