# Treatment of Environmental Data

## Application to UNSD Questionnaire on Environment Statistics

**Wafa ABOUL HOSN**

**Sectoral Statistics Team**

**Sustainable Development and Productivity Division**

**ESCWA**

**Beirut- Lebanon**

**Email: aboulhosn@un.org**

# Outline

1. Introduction
2. Flow Diagram for Environmental Statistics
3. Collecting - Compiling –Treating Data
4. Why is Data Treatment Important?
5. What is the Purpose of the Treatment?
6. Flow Diagram from Data to Information
7. Tools for Data Treatment in Microsoft Excel
8. Level 1. Descriptive Statistics 1: Summary Tables Application to Questionnaire
9. Level 1. Descriptive Statistics 2: Graphs
10. Level 2. Inferential Statistics 1 Application to Questionnaire
11. Level 2. Inferential Statistics 2: Regression- Application to Questionnaire
12. Level 2. Inferential Statistics 3. Forecasting
13. Level 3. Aggregated Data: Indicators, Indices Application to Questionnaire
14. Data treatment Statistics Software
15. Role of ESCWA in Environment Statistics
16. Conclusion

# Introduction

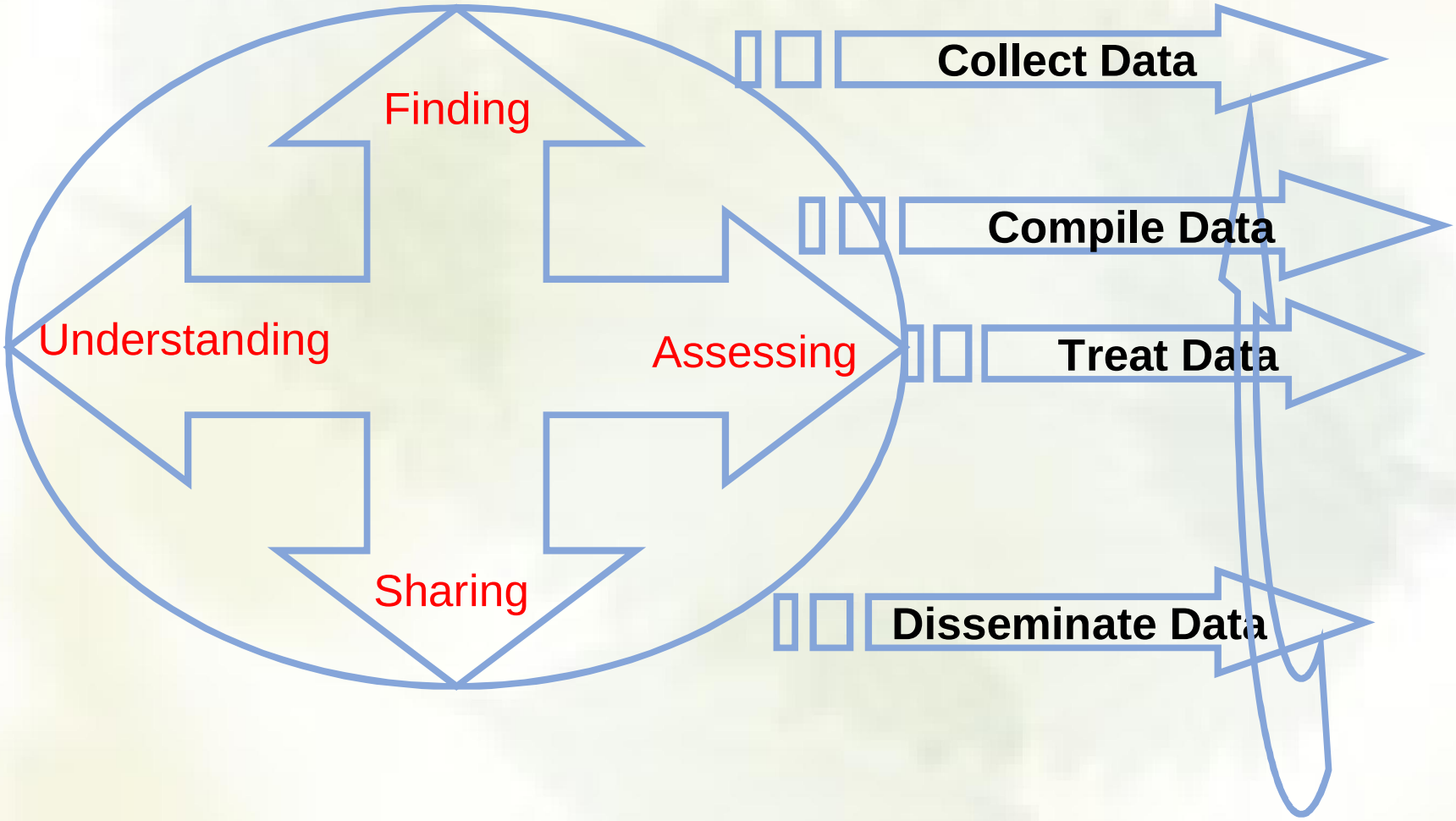
**High quality environmental information is essential for carrying out more responsive and cost-effective policies and providing the right information for the right purpose. It implies:**

**-setting data collection and compiling systems**

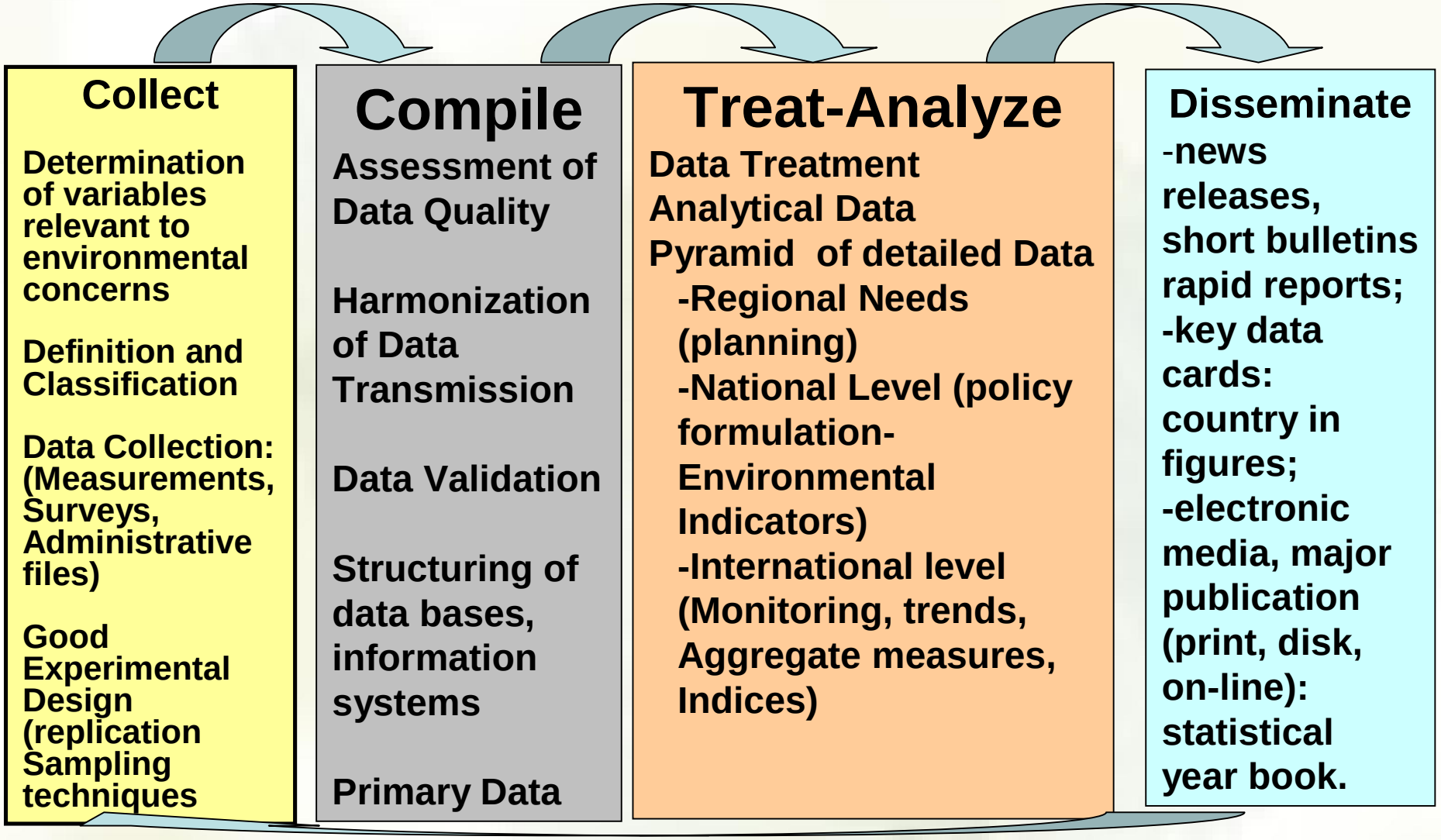
**-ensuring appropriate analysis**

**-providing well targeted dissemination.**

# Environmental Data Components



# Flow Diagram for Environmental Statistics




Cross-Cutting Issues: Multidisciplinarity ,Complexity, Assumptions, Availability, Periodicity, Nature: Quantitative ,Qualitative, GIS, time series

# Collect Data

- **Determination of variables relevant to environmental concerns:**
  - **List of variables in the UNSD questionnaire cover core information concerning environment in the country.**
  - **Integrate other Parameters specific to the ESCWA region (desertification, water scarcity, etc…)**
- **Definition and Classification**
  - **The worksheets Guidance and Definitions of the questionnaire provide definitions of the variables and a conversion table**
- **Data Collection: Measurements: Good Experimental Design**
  **Replication (# of sites)**
  **Sampling techniques (random, stratified, uniform)**
  **Design uniform sampling forms to be used by technicians in the field**

  **Surveys: regrouping or selection before use for environmental issue; Add-on surveys**

  **Administrative files: registers, licenses, …**

# Compile Data

**Assessment of Data Quality**

**Data Validation: Application to water questionnaire**

**Harmonizaton of Data Transmission**

**Structuring of data bases, information systems**

**Primary Data**

## Methods for Validating Data Example: Water Questionnaire: Precipitation:

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Section: WATER** | | | | | | | | | |
| | | | | | | | | | |
| **Country: Belarus** | | | | Contact person: | | | | | |
| Contact institution: | | | | E-mail: | | | | | |
| | | | | | | | | | |
| **Table W1: Renewable Fresh Water Resources** | | | | | | | | | |
| | | | | | | | | | |
| **Priority** | **Category** | **Unit** | **Long term annual average** | **1990*** | **1995*** | **1996** | **1997** | **1998** | **1999** |
| ! | Precipitation (1) | mio $m^3/y$ | 136185.6 | 151963.2 | 127258.8 | 125805.6 | 138054 | 172045 | 116286 |

**Steps to follow in Excel to Validate:**

1. From menu choose Data
2. Select Validation:
3. Data Validation:
4. Allow: for example decimals
5. Data: between
6. Minimum: Choose a minimum
7. Maximum: Choose a maximum
8. Input message
9. Error Alert

# Why is Data Treatment Important?

## Treat-Analyze

**Data Treatment**
**Analytical Data Pyramid of detailed Data**

- **-Regional Needs (planning)**
- **-National Level (policy formulation-Environmental Indicators)**
- **-International level (Monitoring, trends, Aggregate measures, Indices)**

**In General**

- **To better understand data and improve its quality**
- **To add value to data and improve its usefulness**
- **To get closer to analysts and understand their needs**
- **To provide objective and impartial analyses**

**For Environmental Data**

- **To monitor environmental conditions and trends**
- **To promote international harmonization of data**

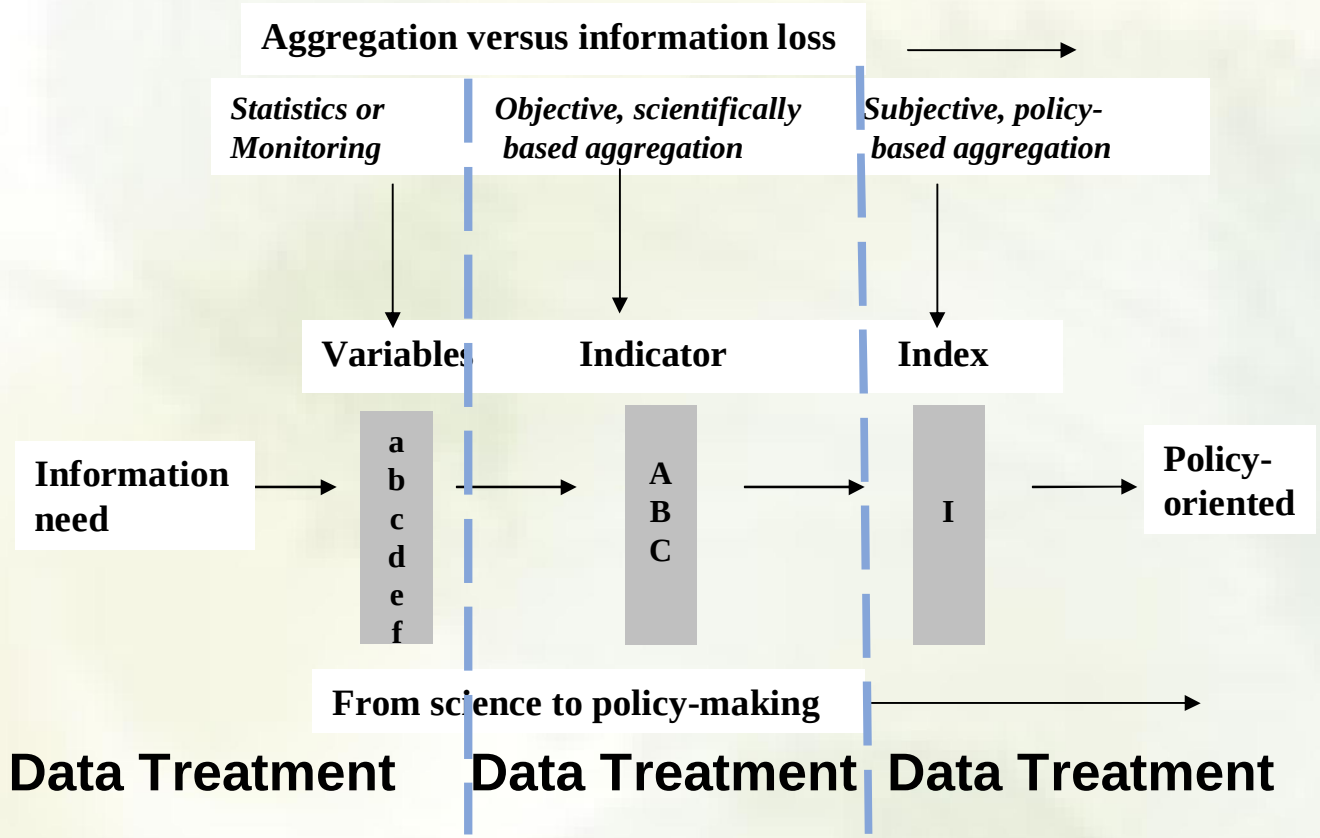
# What is the Purpose of the Treatment?

- ü To Explain
- ü To Report
- To Explore and apply new analytical techniques
- To Improve methods for producing statistics
- To Develop new statistical products

***In the context of this workshop, we shall concentrate on the first two purposes***

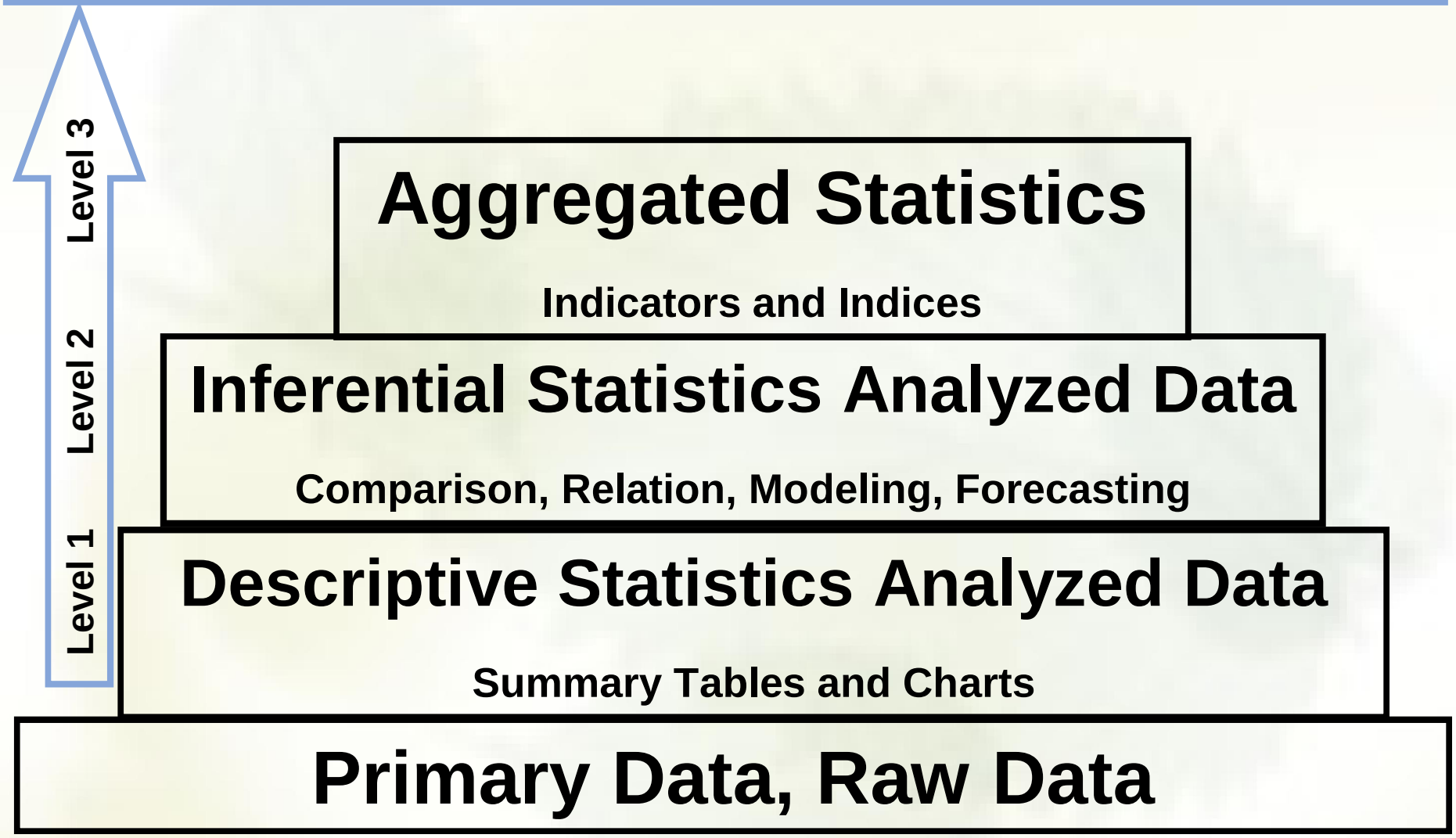
# Flow Diagram from Data to Information

**Figure 1. Translation of an information need into policy-oriented information using variables, indicators and indices**




**This figure shows the difference between variables, indicators and indices, which all represent different stages of information collation. Indicators take variables and condense them into manageable information sets, which are then further condensed by indices. These can then be translated into policy-oriented information. Source: Lorenz, 1999**

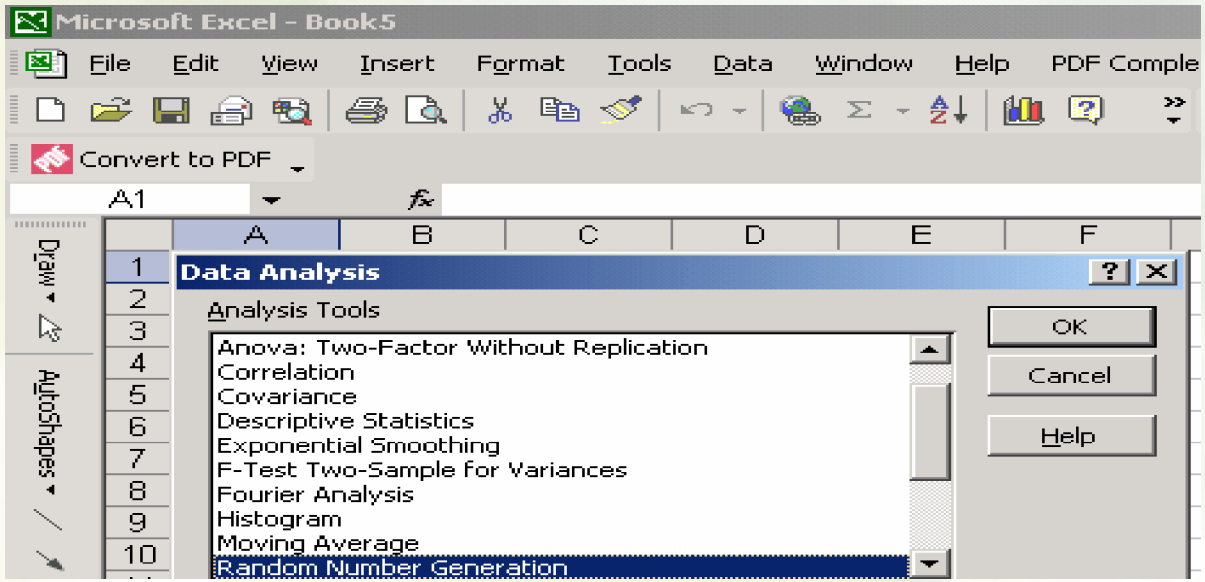
# Statistics from Data to Information



# Tools for Data Treatment in Microsoft Excel

- ***To install the tools for data treatment in Excel:***

1. ***Select Tools from menu***
2. ***Choose Add-Ins…***
3. ***Check Analysis ToolPack***
4. ***Ok***



# Level I. Descriptive Statistics 1: Summary Tables

**All data can be summarized in tables (incorporated into a final report, either in the body of the report (when the number of variables is small), or as appendices. Summary tables should be compiled and include basic statistics (# of values, minimum, maximum, mean, standard deviation, and period of record)**

**AMBIENT WATER QUALITY DATA SUMMARY OF HYPOTHETICAL SITE FOR THE PERIOD OF 1991-92**

| CHARACTERISTICS | # OF VALUES | MINIMUM | MAXIMUM | MEAN | STD DEV. |
|---|---|---|---|---|---|
| GENERAL | | | | | |
| Acidity T4.5 (mg/L) | 13 | 30.9 | 34.1 | 32.6923 | 1.23184 |
| Acidity P8.3 (mg/L) | 13 | <0.5 | <0.5 | <0.5 | 0.0 |
| Coliform (CFU/cL) | 14 | 1 | 2 | 1.9285 | 0.26726 |
| Color (true) (col. units) | 1 | <5 | <5 | <5 | |
| Chlorophyll *a* (µg/L) | 14 | 0.6 | 3.3 | 1.9071 | 0.89224 |
| Dissolved oxygen (mg/L) | 229 | 4.13 | 15.56 | 11.3185 | 1.827 |
| pH (pH units) | 237 | 5.95 | 7.9 | 6.90101 | 0.42637 |
| Secchi depth (m) | 15 | 8.7 | 16 | 13.0433 | 1.93362 |
| Specific cond (µS/cm) | 9 | 72 | 79 | 74.555 | 2.408 |
| Temperature (°C) | 234 | 4 | 25 | 9.47829 | 5.99845 |
| Turbidity (NTU) | 17 | 0.1 | 0.9 | 0.32941 | 0.20237 |

# Application to the Water Questionnaire

**Type of Analysis: Descriptive Statistics**

**STEPS TO FOLLOW:**

1. Arrange data to be congruent
2. Select from menu Tools Data Analysis
3. Select Descriptive Statistics
4. Select for Input range the cells that contain data
5. Check Summary Statistics
6. Output range Select the area whee you want the output to be displayed

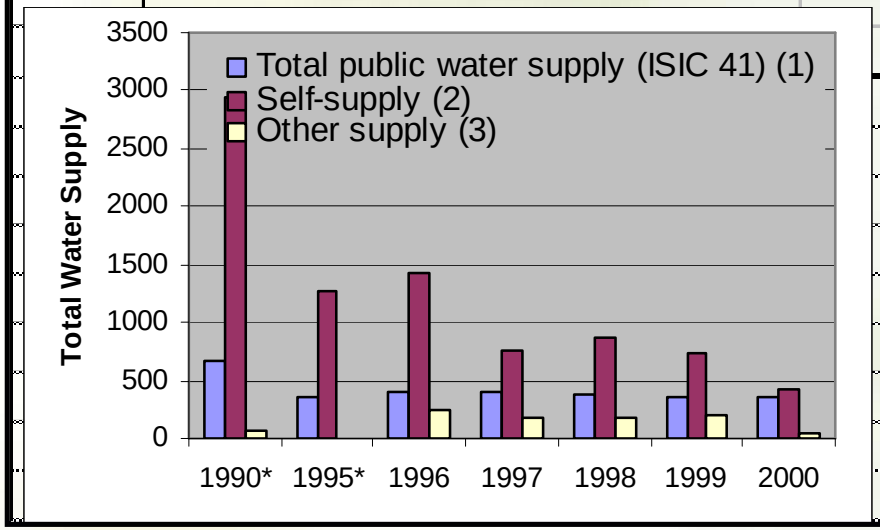
| | | | | | |
|---|---|---|---|---|---|
| Total Precipitation | 127258.8 | 125805.6 | 138054 | 172045 | 116286 |
| Total renewable fresh water resources (5)=(3+4) | 51764 | 70860.8 | 65372 | 122389 | 71999 |

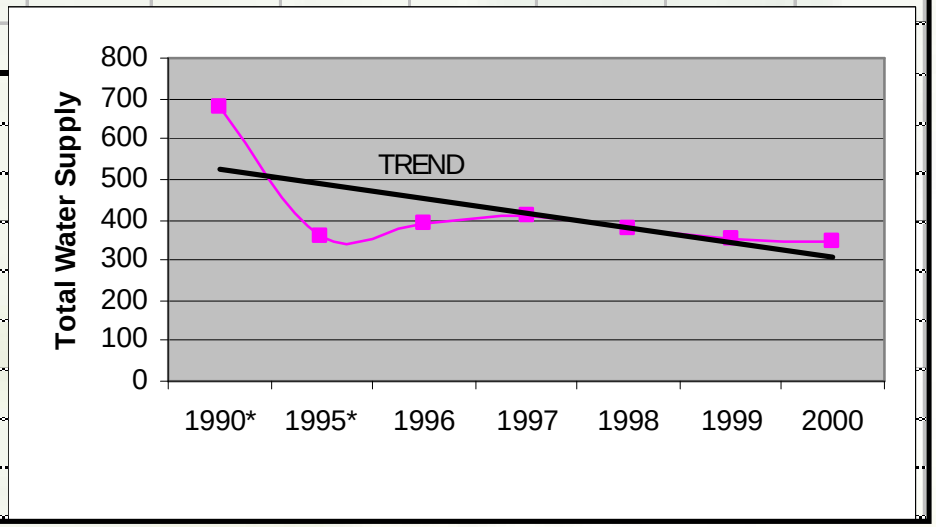
| *Summary Stats 1995-2000* | Precipitation (1) | Actual evapotranspiration (2) | Internal flow (3)=(1)-(2) | Actual external Inflow of surface and ground waters (4) | Total renewable fresh water resources (5)=(3+4) | Outflow of surface and ground waters |
|---|---|---|---|---|---|---|
| | 138198.657 | 84830.8 | 53367.8571 | 26200 | 79567.8571 | 64442.8571 |
| **Mean** | 7059.98565 | 3764.7699 | 6651.49377 | 2546.8935 | 8571.25133 | 6517.69961 |
| **Standard Error** | 135978 | 86456 | 53560.8 | 27600 | 71999 | 65700 |
| Median | 18678.9663 | 9960.645 | 17598.1984 | 6738.4469 | 22677.3995 | 17244.2123 |
| **Standard Deviation** | 348903782 | 99214448 | 309696586 | 45406667 | 514264446 | 297362857 |
| Sample Variance | 0.79774521 | -1.6151266 | 1.6312836 | -0.6095157 | 1.70383716 | -0.8875145 |
| **Range** | 0.98945063 | -0.1230756 | 0.76619613 | 0.1980322 | 1.09180397 | 0.29926674 |
| **Minimum** | 55759 | 25950 | 56525 | 19500 | 70625 | 48700 |
| **Maximum** | 116286 | 72244.8 | 29064 | 17300 | 51764 | 42600 |
| Sum | 172045 | 98194.8 | 85589 | 36800 | 122389 | 91300 |
| **Count** | 967390.6 | 593815.6 | 373575 | 183400 | 556975 | 451100 |

# Level 1. Descriptive Statistics 2: Graphs

**Graphical displays serve as an aid in the data presentation and interpretation**

| Type of Analysis: | Descriptive Statistics: Graphical Display | | | | | | |
|---|---|---|---|---|---|---|---|
| | **1990*** | **1995*** | **1996** | **1997** | **1998** | **1999** | **2000** |
| Total public w ater supply (ISIC 41) (1) | 676.8 | 361.4 | 392.4 | 411.9 | 378 | 355 | 346 |
| Self-supply (2) | 2946.2 | 1266.6 | 1426.6 | 761.1 | 863 | 728 | 432 |
| Other supply (3) | 62.6 | 9.98 | 237 | 173.7 | 174 | 192 | 38 |
| **Total water supply (4) = (1)+(2)+(3)** | 3685.6 | 1628 | 1819 | 1173 | 1241 | 1083 | 778 |
| Midyear Population Estimates (Population in thousan | 5,457 | 5,287 | 5,216 | 5,154 | 5,100 | 5,055 | 5,020 |
| Ref. US Census Bureau | | | | | | | |



# Level 2. Inferential Statistics 1

**Parametric tests can be used If assumptions for (normality, independent random samples ,equal variances are satisfied, if not, use non-parametric tests -Hypothesis Testing (the F-test and the t-test, ANOVA tests) to determine if significant differences exist (either spatially or temporally).**

**- Correlation Analysis**

STEPS TO FOLLOW

1 Arrange data to be congruent
2 Select from menu Tools Data Analysis
3 Select Correlation
4 Select for Input range the cells that contain data
5 Output range Select the area whee you want the output to be displayed

| | Total gross fresh water abstraction | Total Precipitation | Total renewable fresh water resources |
|---|---|---|---|
| Total gross fresh | 1 | | |
| Total Precipitatio | -0.253952518 | 1 | |
| Total renewable f | -0.407718892 | 0.9758462 | 1 |

**Type of Analysis: Regression**

STEPS TO FOLLOW

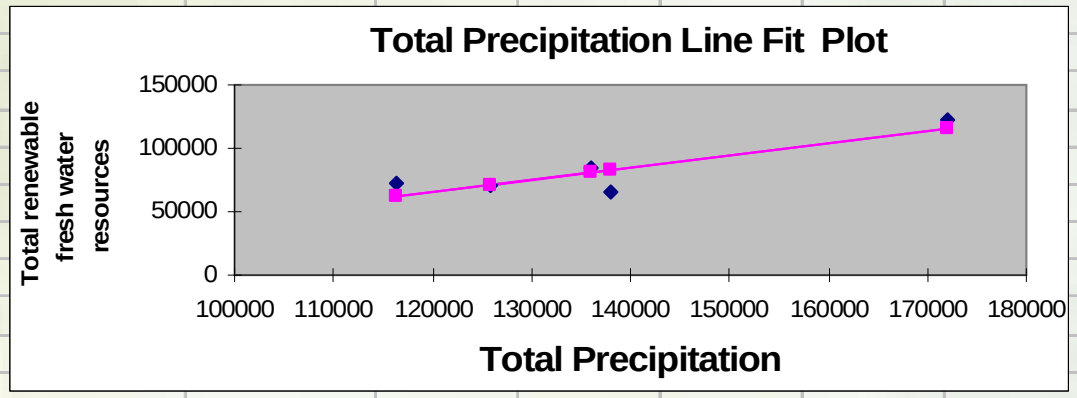
1 Arrange data to be congruent in columns
2 Select from menu Tools Data Analysis
3 Select Regression
4 Select for Input range the cells that contain data for X and Data for Y
5 Check Residuals and line fit

# Level 2. Inferential Statistics 2: Regression

**Type of Analysis: Regression**

**STEPS TO FOLLOW**

1. Arrange data to be congruent in columns
2. Select from menu Tools Data Analysis
3. Select Regression
4. Select for Input range the cells that contain data for X and Data for Y
5. Check Residuals and line fit
6. Output range Select the area whee you want the output to be displayed




SUMMARY OUTPUT

| *Regression Statistics* | |
|---|---|
| Multiple R | 0.885093 |
| R Square | 0.783389 |
| Adjusted R | 0.711185 |
| Standard E | 12404.69 |
| Observatio | 5 |

ANOVA

| | *df* | *SS* | *MS* | *F* | *ignificance F* |
|---|---|---|---|---|---|
| Regressior | 1 | 1669513345 | 1669513345 | 10.84971 | 0.045944 |
| Residual | 3 | 461628996.9 | 153876332.3 | | |
| Total | 4 | 2131142342 | | | |

| | *Coefficients* | *Standard Error* | *t Stat* | *P-value* | *Lower 95%* | *Upper 95%* | *ower 95.0%* | *Jpper 95.0%* |
|---|---|---|---|---|---|---|---|---|
| Intercept | -50315.02 | 40815.2683 | -1.23274992 | 0.305464 | -180207.5 | 79577.5 | -180207.5 | 79577.5 |
| Total Preci | 0.967738 | 0.293797945 | 3.293889565 | 0.045944 | 0.032741 | 1.902735 | 0.032741 | 1.902735 |

# Level 2. Inferential Statistics 3. Forecasting

| Type of Analysis: | Forecasting |
|---|---|

**STEPS TO FOLLOW**

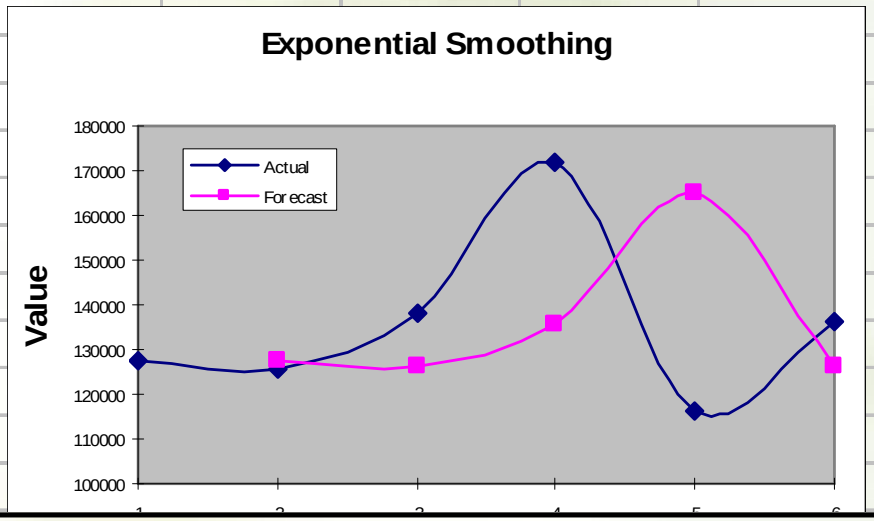
**Forecasting**

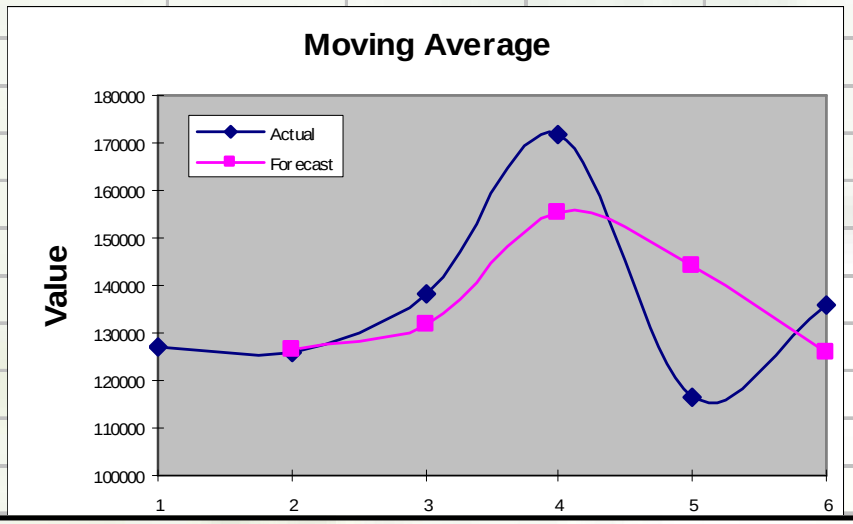
Tools Data Analysis

Exponential Smoothing: Decide how damping factor: 0.2 or 0.3

Moving Average: Decide how many steps back :for example 2

| | | |
|---|---|---|
| 127258.8 | 126241.6 | 134510.3 |
| 126532.2 | 131929.8 | 155049.5 |



# Level 3. Aggregated Statistics: Indices

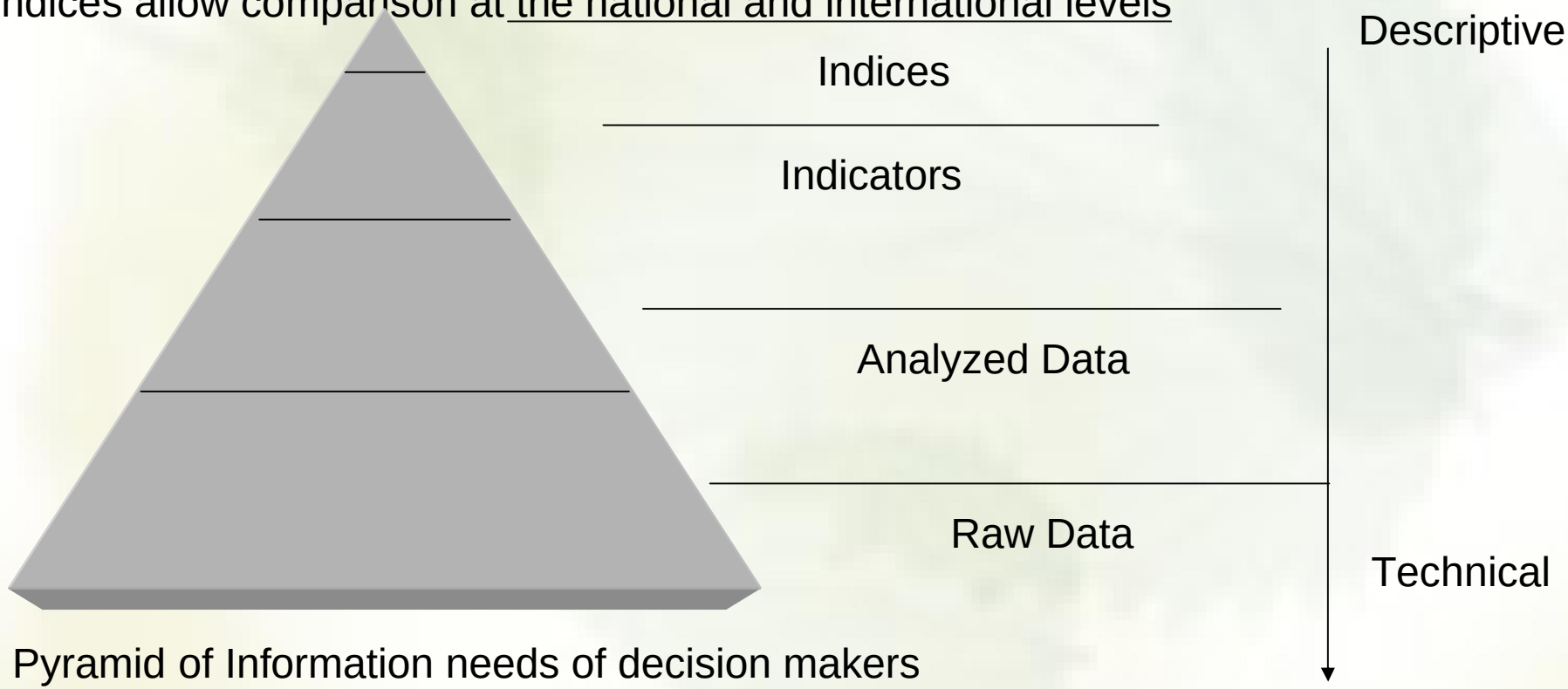
An index is a mathematical means of calculating a **single value from multiple test results.**

Indices are highly **aggregated** Indicators

They **reduce** the number of indicators and add interpretive value to the process

**Weighing** indicators relative to each others is important in designing the index

Indices allow comparison at the national and international levels




Pyramid of Information needs of decision makers

# Importance of Indices

**Importance of Indicators and indices of Environment in the ESCWA**

**-To assist decision makers at the national level in monitoring progress towards sustainable development**

**-To provide scientists with a tool of understanding of human effects on the environment**

**-To support citizens in defining their remedial action and plan for lobbying campaigns**

**-To harmonize at the regional and global levels and compare uniform indicators**

**-To facilitate national reporting**

**-To help the international community to study global changes**

**-To provide local authorities with instruments to develop policies and raise public awareness in the communities.**

# Examples of Water Indices

| | Detailed information | Aggregated Information |
|---|---|---|
| Indicators of Use | Annual extraction per capita (m3) | Water Vulnerability Index |
| | Annual extraction by sector (%) | |
| Indicators of Demand | Total demand (m3) | |
| | Use efficiency (%) | |
| | Recycling potential (%) | |
| Indicators of Generation | Number of dams (no) | |
| | Kilowatts per hectare inundated (kW) | |
| | Hydroelectricity production (mW) | |
| Indicators of Emissions | N emissions (kg) | |
| | Other emissions (kg) | |
| Indicators of Availability | Reserves (m3) | Water Quality Index |
| | Rate of recharge (m3 yr-1) | |
| | Annual rainfall (mm) | |
| | Annual extraction as % of total (%) | |
| Indicators of Quality | Biological oxygen demand(mg L-1) | |
| | Chemical oxygen demand (mg L-1) | |
| | Eutrophication | |
| | Acidification | |
| | Colibacilli (m L-1) | |
| Indicators of Effects | People affected by diarrheic diseases (#) | Climatic Risk Index |
| | Population affected by inundation (#) | |
| | Toxicity/ Heavy metal concentration | |
| Indicators of Risk | Population risking inundations (no) | |
| | Capital risking inundations ($) | |
| Indicators of Protection | Watershed land use | Safe Water Index |
| | Watershed protected area | |
| Indicators of Satisfaction | Access to potable water (%) | |
| | Access to drains (%) | |
| | Aqueducts (#) | |
| | Treatment of used waters (%) | |
| | Water price (US/m3) | |

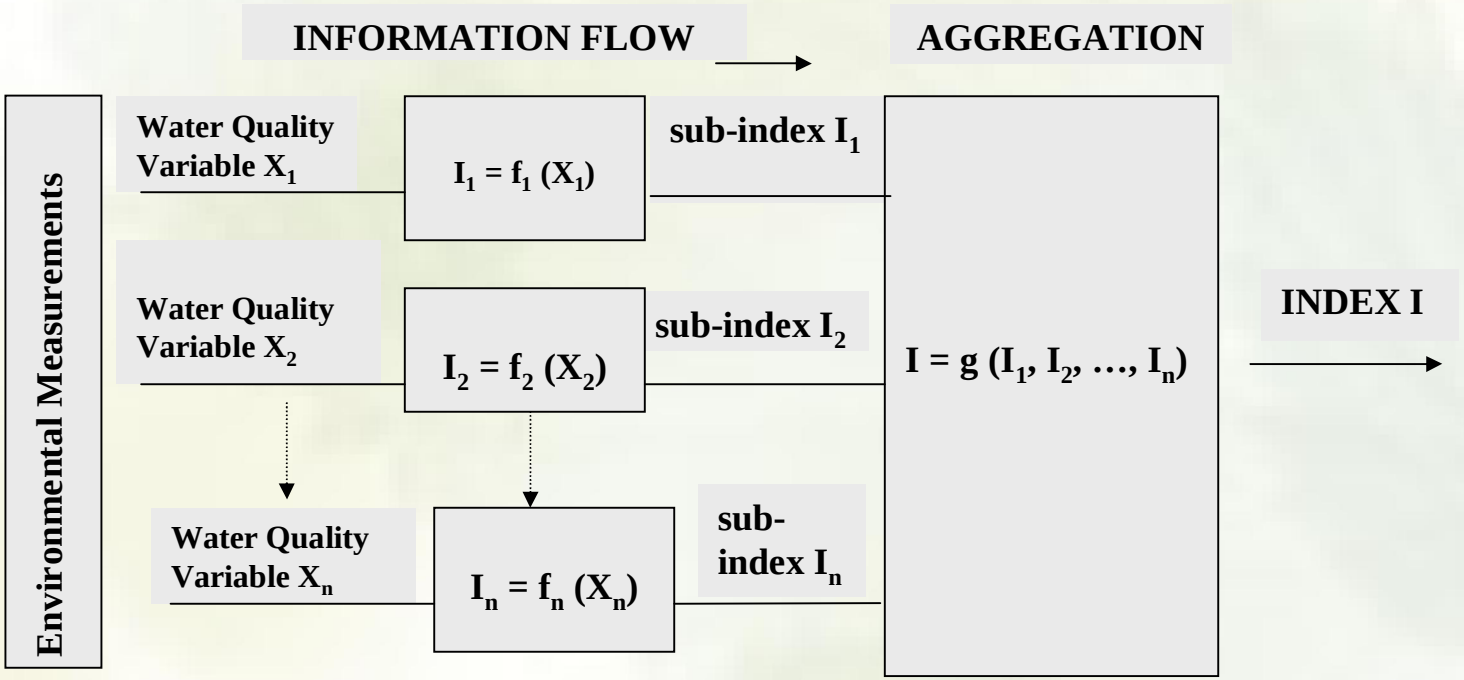
Other indices

Water Stress Index

Water Use Index

Source: CIAT/World Bank/UNEP Project Rural Sustainability Indicators: Outlook for Central America **Technical Note Conceptual Framework to Develop and Use Water Indicators** by Manuel Winograd, Marta Aguilar, Andrew Farrow (CIAT), Lisa Segnestam, Michael Linddal, John Dixon (World Bank) August,1999, CIAT, Cali, Colombia

# Information flow process in the development of water quality indices



# Calculation of Water Stress Index

| Category | 1990* | 1995* | 1996 | 1997 | 1998 | 1999 | 2000 |
|---|---|---|---|---|---|---|---|
| Total fresh surface water abstracted (1) | 1673 | 876 | 827 | 770 | 759 | 756 | 755 |
| Total fresh ground water abstracted (2) | 1210 | 1104 | 1093 | 1089 | 1092 | 1095 | 1082 |
| Total gross fresh water abstraction (3)=(1)+(2) | 2883 | 1980 | 1920 | 1859 | 1851 | 1851 | 1837 |
| Total renewable fresh water resources (5)=(3+4) | 60619 | 29064 | 53561 | 45672 | 85589 | 42899 | 56171 |
| Midyear Population Estimates (Population in thousands) | 10,215 | 10,404 | 10,409 | 10,404 | 10,394 | 10,382 | 10,367 |
| Annual per capita share of renewable water resources m | 5934.3 | 2793.5 | 5145.6 | 4389.9 | 8234.5 | 4132.1 | 5418.3 |
| Water stress index | 4.75592 | 6.81255 | 3.58471 | 4.07033 | 2.16266 | 4.31479 | 3.27037 |

* http://www.census.gov/

# Data treatment Statistics Software

## Statistical Statistics software for Microsoft Excel

**includes over 30 parametric & non-parametric statistical functions, including multiple linear regression analysis, ANOVA, & chi-square statistics http://www.analyse-it.com/**

## XLStatistics Home Page - Excel, Statistics, Data Analysis

**Contains most standard analyses, analyses using only summary data, power / sample size , nonparametrics, curve fitting , non-linear regression, analysis for 2x2 tables. XLStatistics is not an Excel add-in and all the working and code is visible. A free version for analysis of 1- and 2-variable data is available. http://www.deakin.edu.au/~rodneyc/XLSTATS.HTM**

## MINITAB Release 14 for Windows® Demo

**http://www.minitab.com/products/minitab/14/demo/**

# Role of ESCWA in Environment Statistics

- Advisory Services
- Administration: New team for Sectoral Statistics in the Sustainable Development Division to promote statistics for the environment, water, energy, productivity, agriculture and technology in the ESCWA
- Methodologies:
  - Development of ESCWA's Water Pollution Index (WPI) 1995
    Presentation by Ms. Aboul Hosn Water Quality Indices
  - Integrated Water Resources Management (IWRM) training material for the ESCWA region 2004
- Papers from workshops: Adequacy and Validity of Environmental data in the Arab Region by Hanan Atallah, Water team, SDPD, ESCWA
- On Data Dissemination: Presentation by M. A. Dawachi
  http://unstats.un.org/unsd/methods/statorg/Workshops/Doha/Session2_1_IT_Dewachi_slides_English.pdf
- Publications

الإنتاج والطاقة والمياه والبيئة في منطقة الإسكوا: مؤشرات إحصائية

Distr. GENERAL E/ESCWA/SDPD/2003/3 23 May 2003ORIGINAL: ARABIC

- تقرير عن المهمة الاستشارية الى المكتب المركزي للاحصاء، الجمهورية العربية السورية، "المؤشرات البيئية اللازمة لوضع قاعدة البيانات الاحصائية"، اعداد السيد حسني الخردجي، اللجنة الاقتصادية والاجتماعية لغربي آسيا، وثيقة رقم E/ESCWA/IC/2000/39

- تطور مجموعة من المؤشرات البيئية للاستخدام في التحليل الاقتصادي بمنطقة "مينا"، اعداد السيدابراهيم مصطفى، مجلة دراسات الخليج والجزيرة العربية، العدد 102، السنة 27، يوليو/أغسطس/سبتمبر 2001

# Future Directions

1. The collaboration among ESCWA member countries, Secretariat of ESCWA, UNSD and UNEP is valuable
2. Filling the questionnaire on environment is crucial to improve the quality and quantity of environment statistics in the region
3. Following-up on the completion of the questionnaire and transferring to concerned parties is a common responsibility
4. The appropriate data treatment will help monitor environmental conditions and assess status and trends in the ESCWA region
5. Following standards in data collection, treatment and dissemination will promote harmonization of data among ESCWA countries and at the International level

# References


**A FRAMEWORK FOR THE DEVELOPMENT OF ENVIRONMENT STATISTICS 1984. UN STIESA/STAT/SEA.M/78**

***Development of Water Quality Index* ESCWA's WPI 1995: EGM on the implication of Agenda 21 for Integrated Water Management in the ESCWA region. Amman 2-5 October. 1995**

***Guidelines for Interpreting Water Quality Data. 1998. Ministry of Environment, Lands and Parks Land Data BC, Geographic Data BC for the Land Use Task Force Resources Inventory Committee***
**http://srmwww.gov.bc.ca/risc/pubs/aquatic/interp/interp3.htm**