To cite this / Pentru citare: Petrisor AI (2012), *Urban Statistics and Environmental Analysis and Data Analysis Studio. Course synthesis*, available at http://www.environmetrics.ro/Studenti/Sinteza_SUAM-AAD_EN.pdf

Urban Statistics and Environmental Analysis and Data Analysis Studio

Alexandru-Ionuț PETRIȘOR, PhD (Ecology), PhD (Geography)

INFORMATION ON THE COURSE AND STUDIO

Program of studies	Faculty of Urbanism
Type of course	Required course and studio
Level of course module/unit	BA / 2nd year, 2nd semester
Number of ECTS credits	2 + 1 ECTS (theoretical workload for student: 60 + 30 hrs)
Hours / week	2c+1s
Competences to be developed	I. ***Competences specific to urban and spatial planning*** 1) Understanding urban and territorial development from the ecological and sustainable development standpoints, including environmental impact 2) Abilities for multidisciplinary and trans-disciplinary cooperation, communication and establishment of partnership with all territorial and urban actors 3) Capability to elaborate urban and spatial planning documents and abilities for their critical assessment 4) Understanding economic analyses and different economic conditions for the elaboration of strategies, policies, programs and projects focused on human settlements and territories II. ***Competences specific to statistics and environmental analysis*** 1) Understanding main concepts of statistics and environmental analysis 2) Knowledge of the methods (including computer-assisted) used in statistics and environmental analysis 3) Application of statistics and environmental analysis methods to urban ecosystems 4) Using statistical and environmental information in urban, spatial and landscape planning 5) Interpretation of final results of analyses from the standpoint of their relevance to urban or landscape planning 6) Correct use of the technical jargon and specific concepts
Objectives	Placed among informative disciplines, the course and studio do not aim to train specialist in their curriculum, but instead to familiarize students with the core concepts and methods used in statistical and environmental analysis, including computer-assisted techniques, focusing on their application to urban ecosystems. The assignments and essays will allow for a presentation of the application of methods and use information specific to statistics and environmental analysis in urban, spatial and landscape planning, accenting the interpretation of final results and their relevance for an urban or landscape planner.
Teaching methods	Lectures, PowerPoint presentations, computer demonstration (ArcView, Excel), discussions
Evaluation	**Course** – **4** p. presence and activity + **6** p. examination, **studio** – **4** p. presence and activity + **3** p. assignment + **3** p. – The *presence* and *activity* will be graded as following: o *Presence*: the score is proportional with the total number of meetings attended (students attending all meetings receive **4** points). Absences justified by medical certificate presented no later than the exam or serious personal problems reported in due time, if possible, are not penalized. o The *activity* will be graded comparatively based on the accumulation of points when answering correctly questions asked during the course or studio. The student with the maximum total score will receive **4** activity points, and the one with the least, **0** points. The others will be graded proportionally with the total score. *The presence and activity score* is the average of the scores for presence (max. **4** points) and activity (max. **4** points).

	– *The assignment* will consist of using statistical techniques to analyze a subset of the ***Territorial Statistics for Romania, 2002***. Each student will have a different assignment (requirements or dataset). – *The essay* or *presentation* will consist of using environmental analysis to describe main environmental issues and possible solutions in a region (urban or rural) chosen by the student. Each student can opt out for writing an essay (max. 4 A4 pages) or delivering an oral presentation. – ***Penalties***: Indiscipline will be punished individually and collectively. Undisciplined students will be penalized proportional to the seriousness of their undisciplined behavior with "negative activity points", which will be subtracted from their activity score. Negative activity scores will not be considered in the computation of the score reflecting the minimum activity; the minimum limit will be, in this case, zero. Consequently, a student accumulating at least 7.02 "negative activity points" will have the final score [(-7.02+4)/2]+6=4.49 and will be declared failed. Each third indiscipline case will result into the canceling of bonuses for the whole class. For the studio, additional penalties are due to home-works or essays passing deadlines (loss of all available points) and loss or damage of the printed assignment received from the instructor (loss of all available points plus three points deducted from the total score).
Bibliography	1) Botnariuc N., Vădineanu A. (1982), *Ecologie (Ecology)* [in Romanian], Editura Didactică şi Pedagogică, Bucharest, 438 pp. 2) Dragomirescu L. (1998), *Biostatistică pentru începători (Biostatistics for Dummies)* [in Romanian], Editura Constelații, Bucharest, ISBN 973-97950-8-0, 220 pp.; (alternative) Dragomirescu L., Drane J. W. (2001), *Biostatistică pentru începători (Biostatistics for Dummies)*, Vol. I. *Biostatistică descriptivă (Descriptive Biostatistics)* [in Romanian], 2nd Edition, Editura Ars Docendi, Bucharest, ISBN 973-8118-82-4, 165 pp.; Dragomirescu L., Drane J. W. (2005), *Biostatistică pentru începători (Biostatistics for Dummies)*, Vol. I. *Biostatistică descriptivă (Descriptive Biostatistics)* [in Romanian], 2nd Edition, Editura CREDIS, Bucharest, ISBN 973-734-084-1, 165 pp.; Dragomirescu L., Drane J. W. (2006), *Biostatistică pentru începători (Biostatistics for Dummies)*, Vol. I. *Biostatistică descriptivă (Descriptive Biostatistics)* [in Romanian], 3rd Edition, Editura Credis, Bucharest, ISBN 978-973-734-185-3, 206 pp.; Dragomirescu L., Drane J. W. (2007), *Biostatistică pentru începători (Biostatistics for Dummies)*, Vol. I. *Biostatistică descriptivă (Descriptive Biostatistics)* [in Romanian], 4th Edition, Editura Ars Docendi, Bucharest, ISBN 978-973-734-191-4, 186 pp.; Dragomirescu L., Drane J. W. (2008), *Biostatistică pentru începători (Biostatistics for Dummies)*, Vol. I. *Biostatistică descriptivă (Descriptive Biostatistics)* [in Romanian], 5th Edition, Editura Ars Docendi, Bucharest, ISBN 978-973-734-305-5, 206 pp.; Dragomirescu L., Drane J. W. (2009), *Biostatistică pentru începători (Biostatistics for Dummies)*, Vol. I. *Biostatistică descriptivă (Descriptive Biostatistics)* [in Romanian], 6th Edition, Editura Ars Docendi, Bucharest, ISBN 978-973-734-461-8, 207 pp. 3) Petrişor A.-I. (2007), *Elemente de statistică urbană cu aplicații rezolvate în Excel (Elements of Urban Statistics with Applications Resolved Using Excel)* [in Romanian], Editura Universitară „Ion Mincu", Bucharest, ISBN 978-973-7999-86-3, 51 pp. 4) Petrişor A.-I. (2007), *Analiză de mediu cu aplicații în urbanism şi peisagistică (Environmental analysis with applications in urban and landscape planning)* [in Romanian], Editura Universitară „Ion Mincu", Bucharest, ISBN 978-973-7999-85-6, 89 pp. 5) Petrişor A.-I. (2008), *Ecologie urbană, dezvoltare spațială durabilă şi legislație (Urban Ecology, Sustainable Spatial Development and Legislation)* [in Romanian], Editura Fundației România de mâine, Bucharest, ISBN 978-973-163-305-3, 272 pp. 6) Petrişor A.-I. (2010), *Mediul urban: o abordare ecologică (Urban environment: an ecological approach)* [in Romanian], Urbanistique, http://www.urbanistique.ro/mediul-urban-o-abordare-ecologica-dr-alexandru-ionut-petrisor/#more-127 7) Petrişor A.-I. (2011), *Systemic theory applied to ecology, geography and spatial planning*, Lambert Academic Publishing GmbH & Co. KG, Saarbrücken, Germany, ISBN 978-3-8465-0260-0, 172 pp. 8) Vădineanu A. (1998), *Dezvoltarea durabilă (Sustainable Development)*, Vol. I. *Bazele teoretice ale dezvoltării durabile (Theoretical Foundations of Sustainable Development)* [in Romanian], Editura Universității din Bucureşti, Bucharest, ISBN 973-975-256-5, 248 pp. 9) Vădineanu A., Negrei C., Lisievici P. (*coordinators*) (1999), *Dezvoltarea durabilă (Sustainable Development)*, Vol. al II-lea. *Mecanisme şi instrumente (Mechanisms and instruments)* [in Romanian], Editura Universității din Bucureşti, Bucharest, ISBN 973-575-333-2, 348 pp.

Notes on an Introductory Urban Statistics Course

Note: The text represents a synthesis of core methods taught in the courses *Urban Statistics and Environmental Analysis* and *Data Analysis Studio* designed for ERASMUS students. For any other uses, please request the permission of the author.

General concepts of statistics

The purpose of **statistics** is to study a set of observations on some objects of the same nature called ***statistical units***, displaying ***variable characteristics*** (simply ***variables***) susceptible to be *classed*, *ordered* or *measured*. The set is called ***statistical series*** or ***string***.

There are two types of sets:

– ***Populations*** are sets of objects, individuals, phenomena, events, idea, opinions, numbers etc. focusing the interest of researchers. They are large (mostly infinite) and their exhaustive study is impossible or uneconomical.
– ***Samples*** are subsets of the *populations* drawn to obtain information on populations.

Studies made on *populations* by **descriptive statistics** produce certain results, while those made on *samples* by **inductive/inferential statistics** lead to uncertain results. The scientific expression of uncertainty is given in *inductive statistics* by statistical inference. ***Statistical inference*** represents the extrapolation of judgments from samples drawn through specific statistical-mathematical procedures to populations.

The attempt to explain one or more scientific observations is called ***scientific hypothesis***. These hypotheses need to be sustained by data (experiments, observations) and statistics. ***Statistical hypotheses*** are statements concerning one or more *populations* made to check *scientific hypotheses*. A *scientific hypothesis* consists of a **null hypothesis** ("there are no differences") and a **alternative hypothesis** contradicting it and corresponding to the *scientific hypothesis*. After applying a **statistical test**, the *null hypothesis* is rejected when significant differences are detected or not, otherwise. ***Significant differences*** are too large, compared with a chosen ***level of significance***, to be attributed to random fluctuations, but are due to a significant reason, *i.e.*, the *scientific hypothesis*.

Classification of scales: if A and B are two statistical units, and x some variable with the characteristics xA and xB,

– **The nominal scale** makes a distinction: $xA = xB$ or $xA \neq xB$.
– **The ordinal scale** establishes an order. If $xA \neq xB$, then either $xA > xB$ or $xA < xB$.
– **The equal interval scale** provides a measure of the difference. If $xA > xB$, then A is greater than B with $xA - xB$.
– **The equal ratio scale** has in addition a absolute e zero, implicitly a measure of the ratio of two values: A is xA / xB times greater than B.

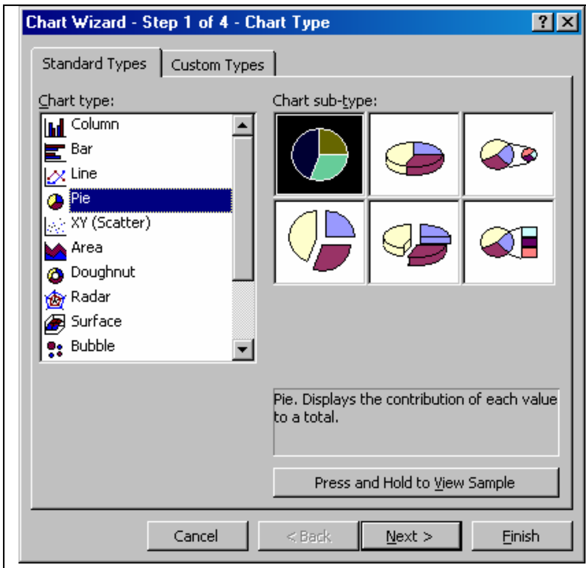
Classification of variables

– Qualitative – nominal scale, including binary variables (yes/no); they have variants that are classed
– Ranks – ordinal scale; they have values that are ordered
– Measures or dimensions – interval or ratio scales; they have values that are measured

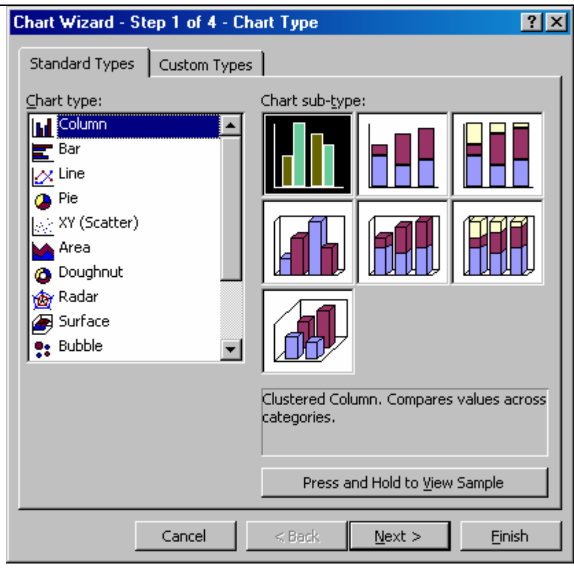
Computing and interpreting results of the graphical synthesis of data

For the assignment, you will receive a data series: X_1, X_2, ..., X_i, ..., X_n. Using this data, you must be able to synthesize them graphically and numerically and interpret the results.

Represent your series as a pie chart (*Fig. 1A*). Write your data in an Excel column, pick the graph sign (), and then choose "pie". You can do it by hand by summing $X_1, X_2, ..., X_i, ..., X_n$. The sum receives an angle 360 degrees. X_1 receives $X_1 * 360 / SUM (X_1, X_2, ..., X_i, ..., X_n)$ degrees etc.



A. Pie chart



B. Bar chart

Fig. 1. Graphical representations using Excel

Represent your series as a bar chart (*Fig. 1B*). Write your data in an Excel column, pick the graph sign (), and then choose "bar". You can do it by hand by summing $X_1, X_2, ..., X_i, ..., X_n$. The sum receives an angle 360 degrees. X_1 receives $X_1 * 360 / SUM (X_1, X_2, ..., X_i, ..., X_n)$ degrees etc.

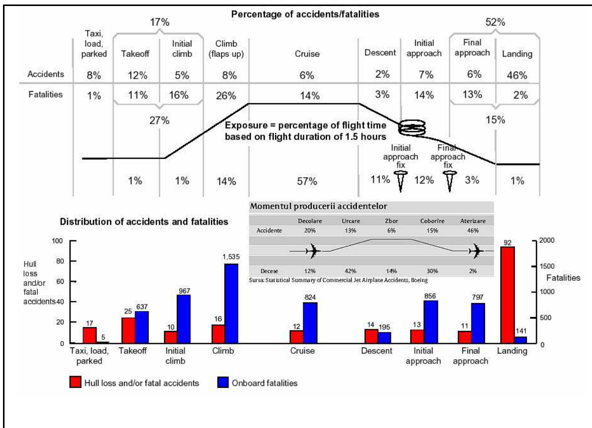
Other graphical representations of data

The following are not required for your assignment. However, they could be useful to know.

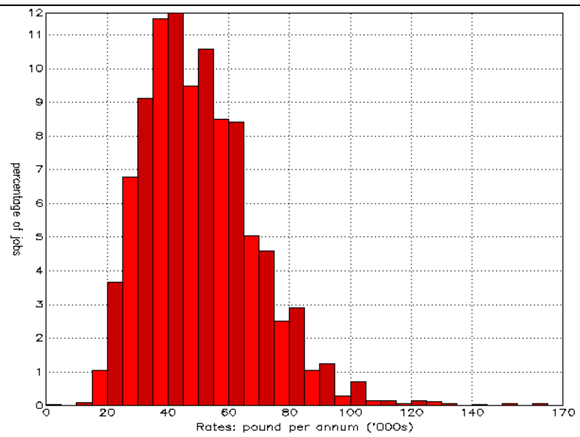
The histogram: unlike the previous ones, data are grouped. A convention requires each bar of a histogram to join the previous one in continuation (unlike the bars of a bar diagram). Histograms are drawn manually or using specialized software. We group data as an essential principle of statistics is to give up information in order to increase the relevance. More concretely, we want to see behind the histogram a theoretical curve, explained below.

Frequency polygons: join upper ends of bars within a bar chart. The line could also suggest some theoretical curve.

Population pyramids: draw two histograms with the age groups for each gender. Rotate them such that their bases are joined. These pyramids suggest the sizes of different population cohorts as a result of historical or environmental pressures.



Bar chart



Histogram

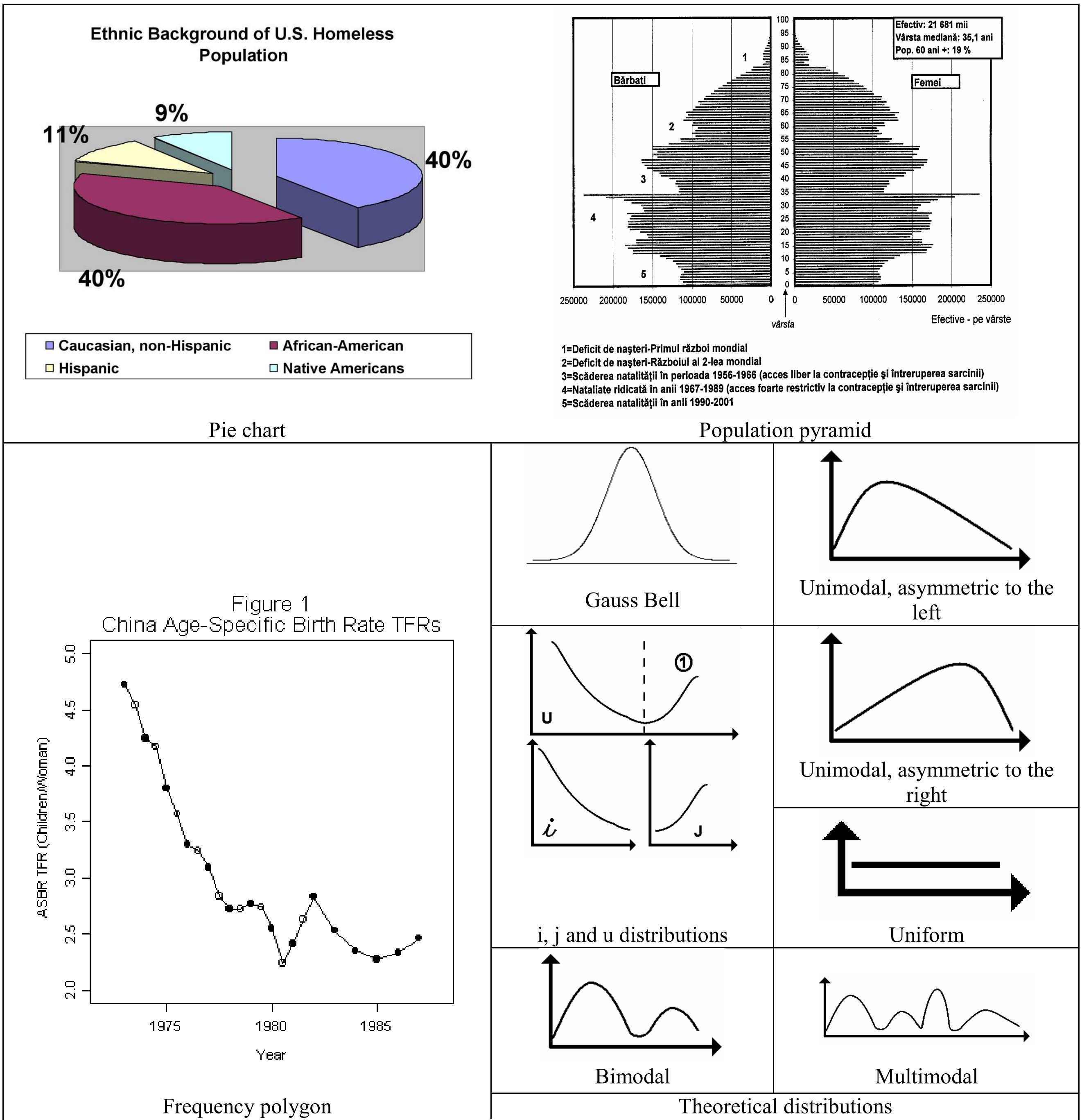


Fig. 2. Representing your data

Theoretical curves: they cannot be drawn directly, but are suggested by histograms and frequency polygons when the sample size increases. Some of them are characteristic to different phenomena and suggest what methods should be used in data analysis. The main representations are summarized in *Fig. 2* below.

Interpreting the theoretical curves

- Distributions concentrated in one point express absolute homogeneity
- Symmetrical distributions are the best to indicate a central tendency
- Bi – or multimodal distributions indicate heterogeneity as a mixture of homogenous distributions
- Uniform distributions express absolute heterogeneity

In statistics, variability is understood as scattering around a central tendency. To understand variability, we must identify the central tendency and analyze how scattered is the distribution around it. This is done using the numerical synthesis of data.

Computing and interpreting results of the numerical synthesis of data

1. Identify the central tendency

Compute M, the average

$$M = \frac{\sum_{i=1}^{n} X_i}{n}$$

Compute Me, the median: this is the value that splits a series into two. Arrange the values in increasing order and either choose the middle one or, if there are two, compute their average (*Fig. 3*).

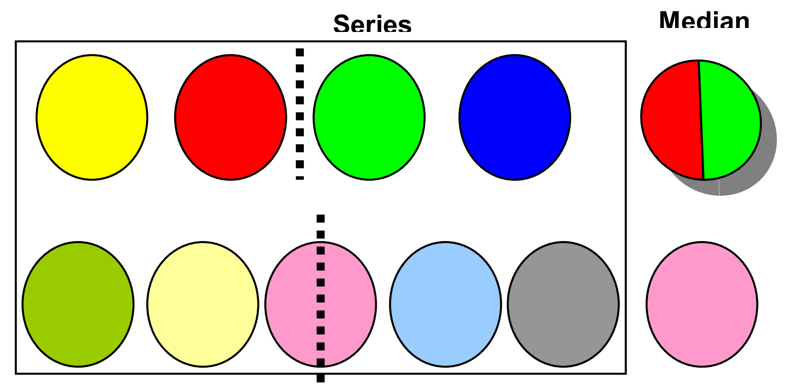


Fig. 3. Computation of median

Compute Mo, the mode and describe the series based on it: this is a value with maximum local frequency (is repeated more times than the values around in a series arranged in increasing order). A series can have one mode (this is called unimodal), two (bimodal) or more (multimodal). If all values have the same frequency (appear the same number of times), the distribution is uniform.

2. Look at scattering (variability)

Compute A, the range: this is the difference between the maximum and the minimum value.

Compute S^2, the variance: subtract from each value the average, square the result (multiply it with itself), add all these results and divide the sum by the number of values. Please note that Excel divides the sum by the number of values minus one, so if you use Excel, you must adjust the result (multiply it with the number of values minus one and divide it by the sum by the number of values).

$$S^2 = \frac{\sum_{i=1}^{n} \left(X_i - M\right)^2}{n}$$

Compute S, the standard deviation: take the square root of the variance

$$S = \sqrt{S^2}$$

Compute CV, the coefficient of variation: divide the standard deviation by the average and multiply with 100 (express it as a percentage). Also, you must classify your series: if CV<10%, your series is homogenous; if 10%<CV<20%, the series is relatively homogenous; if 20%<CV<30%, the series is relatively heterogeneous; if CV>30%, the series is heterogeneous.

$$CV = 100 \times \frac{S}{M}$$

Notes on an Environmental Analysis and Environmental Impact Assessment course

Note: The text represents a synthesis of core methods taught in the courses *Urban Statistics and Environmental Analysis*, *Data Analysis Studio*, *Environmental Analysis and Impact Assessment*, and *Urban Ecology and Impact Environmental Assessment* designed for ERASMUS students. For any other uses, please request the permission of the author.

Environmental Analysis

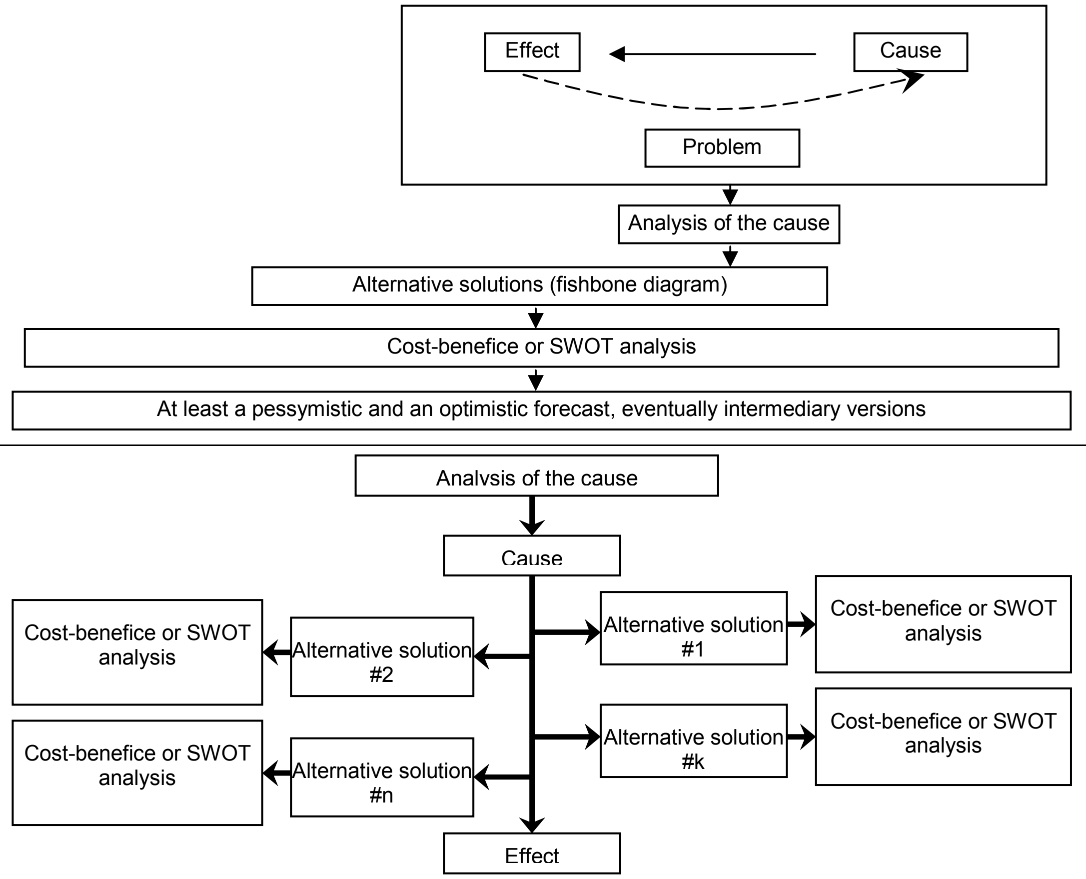


Fig. 1. Analysis and diagnosis of environmental problems (above) using the fishbone diagram (below)

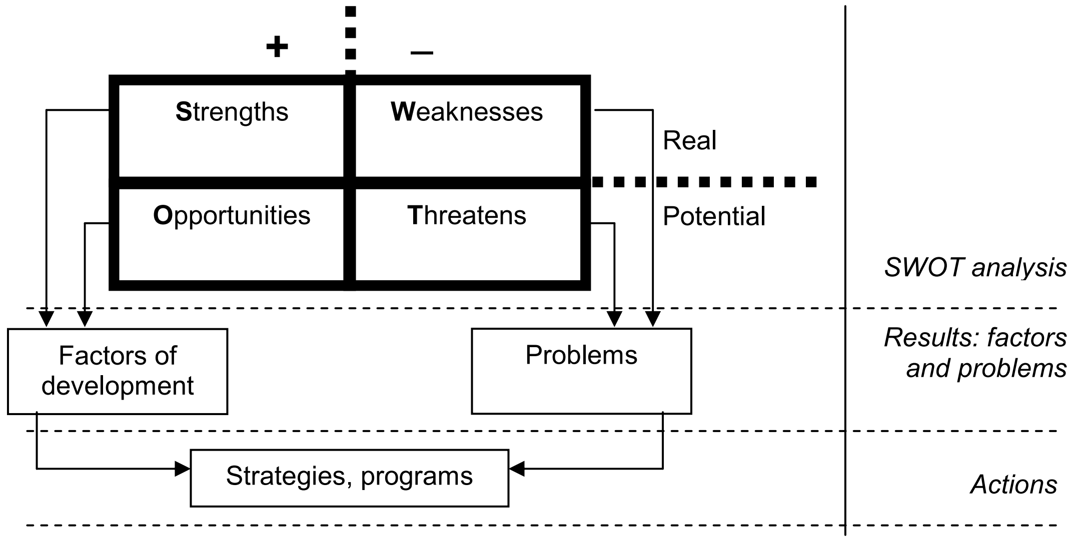


Fig. 2. SWOT analysis

Costs	Benefices
– ...	– ...
– ...	– ...
– ...	– ...

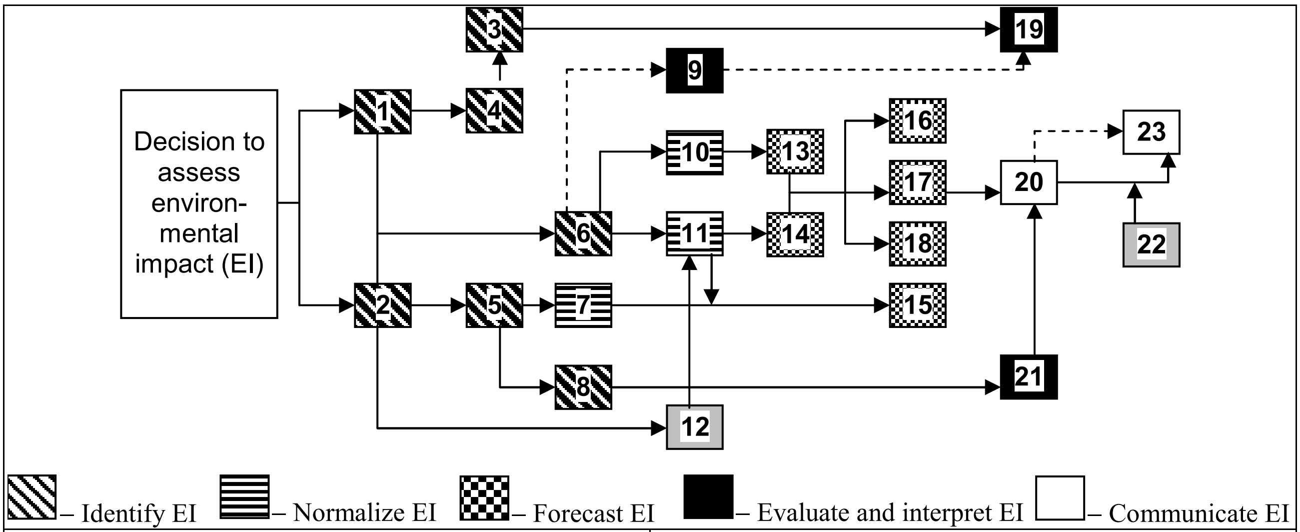
Fig. 3. Cost-benefice analysis

Determining the market value of environmental goods and services

- ***Method of substitution markets***
  - **Environmental protection expenses technique** – individuals would prefer prevention until costs equal availability to pay for decreasing health risks
  - **Hedonistic prices technique** – explain price variations using information on attributes, use statistical regression to estimate price based on the correlation between price and observable attributes
  - **Travel cost technique** – improved environmental quality could increase recreational opportunities in a region – substitute additional satisfaction with value granted to it (money, time)
- ***Hypothetical markets method***
  - **Surveys** – ask people how much would pay for a benefice or to avoid damage cost.
- ***Indirect methods***
  - **"Dose response" or "dose-effect"** – based on the correlation between environmental changes and qualitative/quantitative/financial changes of production.

Table 1. Life cycle of a product

Stage	Idea of product	Design	Technological design	Production	Market	Use – current, capital repairs	Post-use
Time	t_0	t_1	t_2	t_3 $t_{3_1} \rightarrow t_{3_2}$	t_4	t_5	t_6



1 – analyze project and alternatives	13 – reduce impacts
2 – analyze and diagnose area of implementation	14 – assess EI in comparable units
3 – identify potential corrective actions	15 – transform units
4 – identify elements and actions causing EI	16 – assess area if project is applied
5 – identify environmental components to be affected	17 – assess gross impact
6 – identify impact: cause-effect relationships	18 – assess area if project is not applied
7 – select appropriate indices and units	19 – corrective actions
8 – identify corrective actions in the area	20 – final report on EI and communication to the population
9 – qualitative EIA	21 – corrective actions
10 – characterize EI: place, time, manner	22 – participation of population
11 – forecast EI size if the project is not applied	23 – declaration of EI
12 – assess area before implementation	

Fig. 4. Sequences in Environmental Impact Assessment (*apud* Negrei, 1997)

Methods used in Environmental Impact Assessment (EIA)

– 1. **Checklist (Westman, 1985)** – table containing impacts in the lines and phases of the project in columns; mark intersection with X or record value of EI (size, importance)

Example. Checklist (Westman, 1985) applied to a transport project – construction phase

Project	Construction phases
I. Acoustic impact A. Population health B. Land use	**x**
II. Air quality A. Population health B. Land use	**x**
III. Water quality impact A. Underground waters 1. Diffuse surface and underground sources 2. Effects of filling and dragging 3. Characteristics of flooding and draining B. Runoff 1. Effects on effluent loading 2. Other actions: a) Disturb benthos b) Alter currents c) Change flow regime d) Underground salt infiltration 3. Population health 4. Land use	**x**
IV. Soil erosion A. Land use and economy B. Pollution and alluvial deposits	**x**
V. Ecological impacts A. Flora B. Fauna	**x**

– *2. **Cause-effect matrix (Leopold, 1971)** – two-entry table: structure/functions of ecosystem and actions generating impacts*

Example. Phosphate mine construction. First number indicates size, and the second estimates the importance of EI.

	Industrial objectives	Highways, bridges	Communication	Mining	Ground digging	Treatment of minerals	Truck transport	Waste deposits	Runoff, infiltration
Water quality					2/2	1/1		2/2	1/4
Air quality						2/3			
Erosion		2/2			1/1			2/2	
Deposits and sediment						2/2		2/2	
Shrubs						1/1			
Grass						1/1			
Water plants						2/2		2/3	1/4
Fish						2/2		2/2	1/4
Tourism						2/4			
Landscape	2/3	2/1	2/3		3/3		2/1	3/3	
Wildlife quality	4/4	4/4	2/2	1/1	3/3	2/5	3/5	3/5	
Rare species		2/5		5/10	2/4	5/10			
Health and security									

– *3. **Matrix of large dams** – two-entry table similar to the previous, but "inversed", used in conjunction with the **method of networks (Sorensen, 1971)***

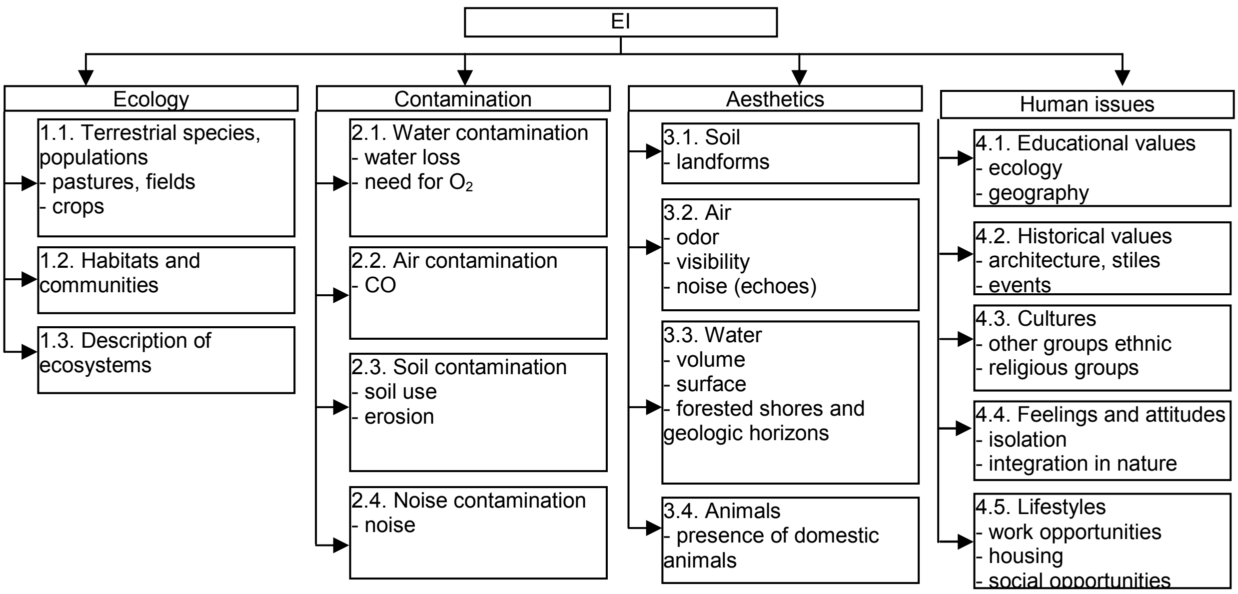
Example. Use of arrows to indicate chain reactions and closed loops

	Agriculture and husbandry	Recreational utilities	Erosion	Sediments	Solid alluvia	Physical and chemical quality of water	Aquatic flora	Aquatic fauna
Irrigation								
Fishing								
Dam								
Anti-erosion curtains								
Artificial lake								
Top layer (live)								
Dragging								
Erosion control								
Leveling pool								

Interpretation. Irrigations favor development of agriculture and husbandry, leading to accented soil erosion and sedimentation, altering physical and chemical quality of water. For this reason, proposed anti-erosion curtains could control the process. Sedimentation is accented close to the dam, but also in the artificial lake, requiring dragging. An alternative solution is a leveling pool. Dragging intensifies alluvial process, and aquatic flora and fauna. Dragging leads to a decrease of the fishing harvest.

– *4. **Battelle-Columbus method**: tree structure of the impact area – categories, components, units, data*

Example.



– *5. **Delphi method** – qualitative method based on input from specialists, data processed using: (1) sorting by branches, (2) scale-based classification or (3) pair comparison.*