Presenting and communicating statistics. Principles, components and assessment

Filomena Maggino

Università degli Studi di Firenze

The study presented here is the result of a project developed by myself and

Marco Fattore
Università degli Studi di Milano-Bicocca
and
Marco Trapani
Università degli Studi di Firenze

Contents

1. Communication: full component of the statistical work
2. Communicating statistics
3. Assessing the quality of communication in statistics

Contents

1. Communication: full component of the statistical work
2. Communicating statistics
3. Assessing the quality of communication in statistics

Communication in statistics: From DATA to MESSAGE

DATA PRODUCTION	→	objective observation		aseptic data
↓				
DATA ANALYSIS, RESULTS AND INTERPRETATION	→	data	transformed in	information
↓				
PRESENTATION	→	information		message

Communication in statistics: From DATA to MESSAGE

data production	data analysis	representation	communication

not only a technical problem

a formula…

VAS= N*[(QSA*MF)*RS*TS*NL] → ***Giovannini, 2008***

This detailed formula, including many relevant aspects like the role of media and users' numeracy, can be reconsidered by including also aspects concerning "quality" e "incisiveness" of the message:

VAS = f (N,QSA,MF,RS,TS,NL,QIP) → ***additional component***

VAS	Value added of official statistics
N	Size of the audience
QSA	Statistical information produced
MF	Role of media
RS	Relevance of the statistical information
TS	Trust in official statistics
NL	Users' "numeracy"
QIP	Quality and incisiveness of presentation

statistics …

… cannot be presented in an aseptic and impartial way by leaving interpretation to the audience

Interpretation …

… can be accomplished through different even if correct perspectives

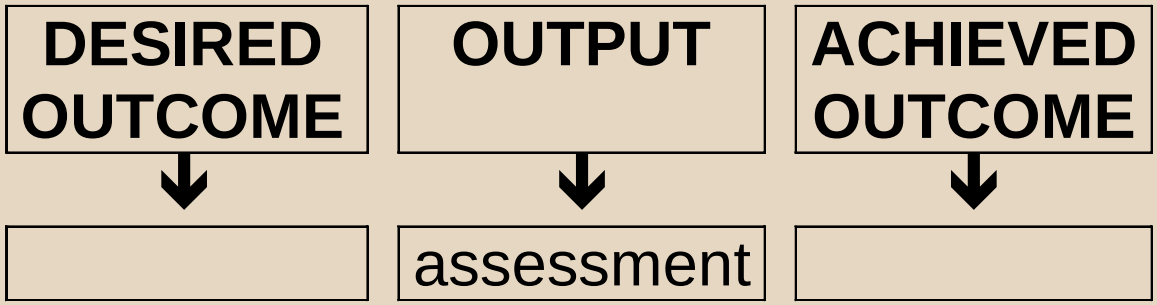
"the glass is half-full" ← → "the glass is half-empy"

through a dynamic perspective

"the glass is getting filled up" ← → "the glass is getting empty"

The message will be transmitted and interpreted by the audience without realizing the mere numeric aspect.

Communication in statistics: from DATA to MESSAGE



statistician

↓

facilitator

between reality and its representation

COMPLEXITY

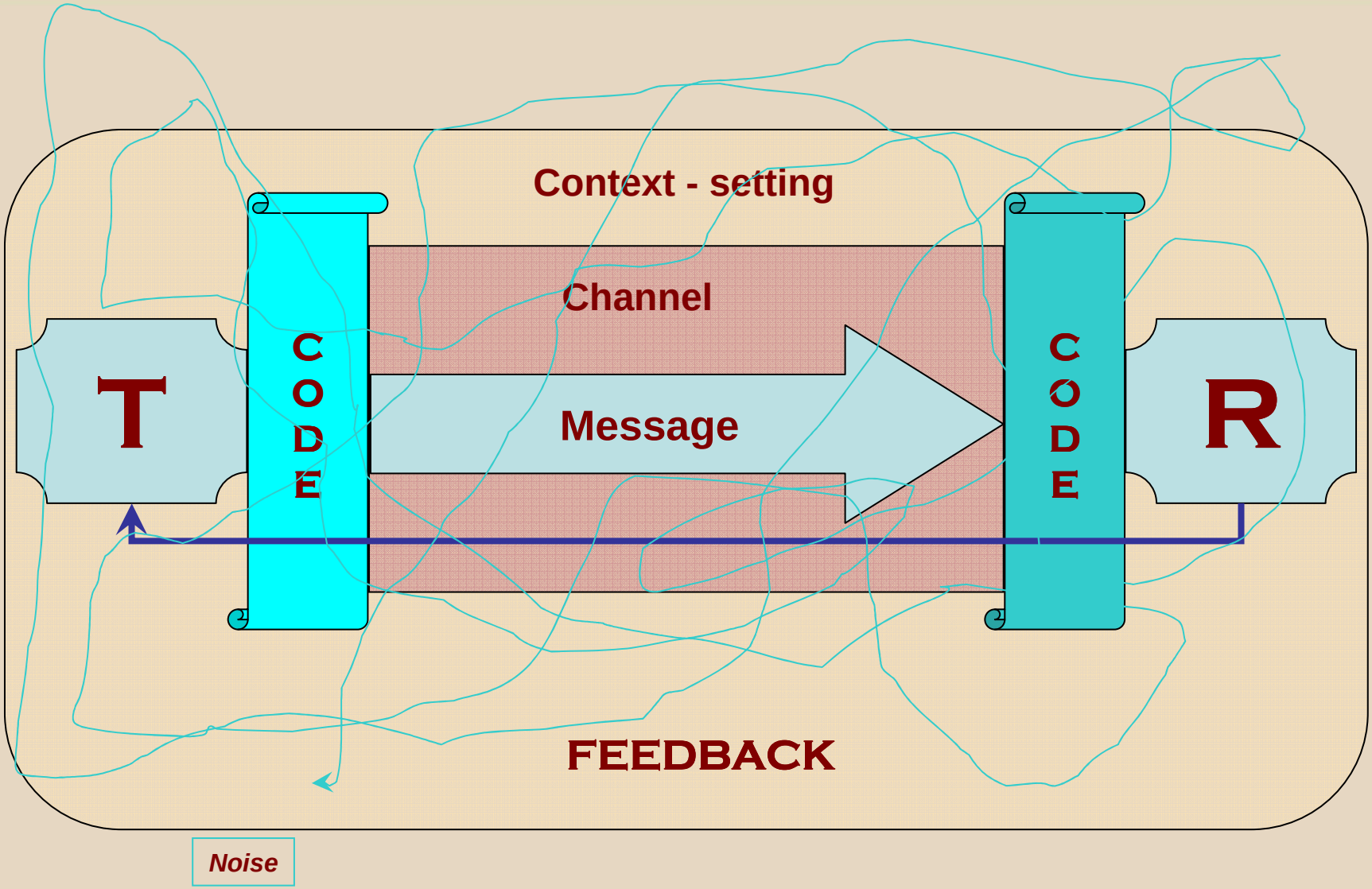
Contents

Contents

1. Fundamental aspects

Aspects of statistical presentations	Corresponding discipline
Content	Ethics
Appeal	Aesthetics
Persuasion	Rhetoric
	↓ **Theory of presentation**

2. Main components

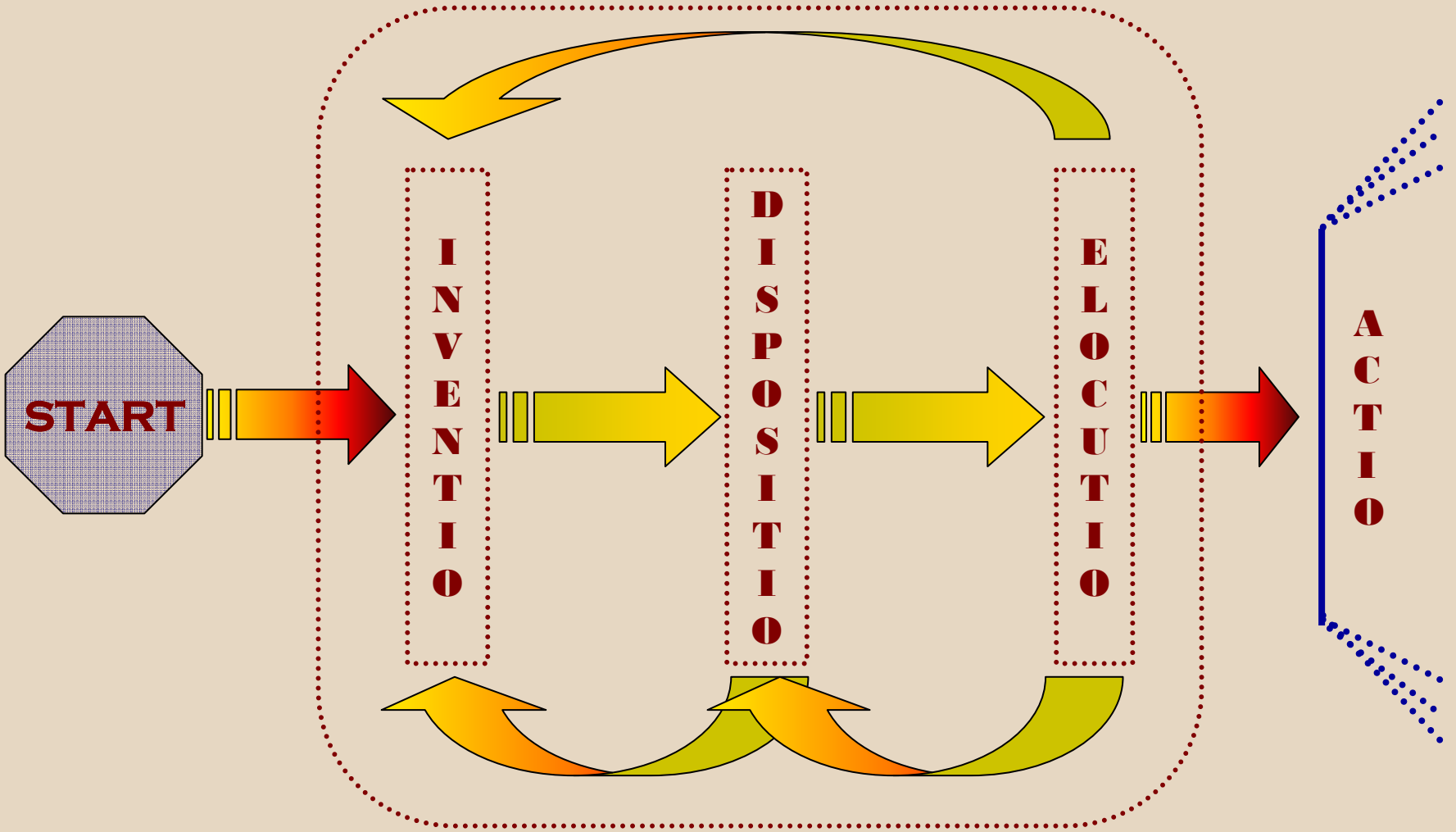


3. Codes

in statistical communication

A. Outline → telling statistics
B. Tools → depicting statistics
C. Clothes → dressing statistics

A. Outline → telling statistics



A. Outline → telling statistics

1- *INVENTIO* (INVENTION)

allows arguments to be argued

Who	→	the subject of telling
What	→	the fact
When	→	the time location
Where	→	the field location
Why	→	the causes

A. Outline → telling statistics

2- *Dispositio* (layout)

allows topics to be put in sequence

- deductive
- inductive
- time-progression
- problems-related

- advantages-disadvantages
- from-points-of-view
- top-down approaches

A. Outline → telling statistics

2- *Dispositio* (layout)

Deductive approach

- Premise
- General Principles
- Developing arguments
- Pratical consequences/examples

Inductive approach

- ***Case / specific situation***
- ***Reflection***
- ***Concepts***
- ***Consequences / other cases***

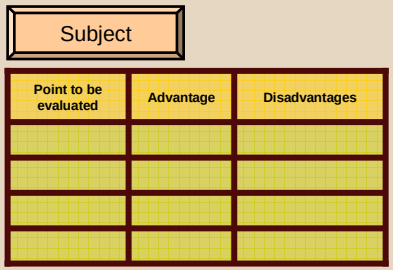
Time progression approach

- ***Once upon a time…***
- ***Why something changed***
- ***Yesterday… Today…***
- ***Tomorrow***

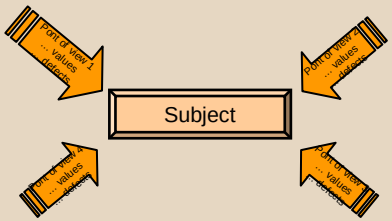
Problems approach

- ***Meaningful questions***
- ***Why in important to talk about…***
- ***Solutions (and concepts)***
- ***Conclusions and consequences***

Advantages-disadvantages approach



From point of view approach



Top-down approach

Premise	Reflections	Concepts	Consequences…
General	Reflections	Concepts	Consequences…
Particular	Reflections	Concepts	Consequences…
Specific	Reflections	Concepts	Consequences…
Detail	Reflections	Concepts	Consequences…
Micro	Reflections	Concepts	Consequences…

A. Outline → telling statistics

3- *Elocutio* (expression)

allows each piece of the presentation to be prepared by selecting words and constructing sentences

Language should be
- appropriate to the audience
- consistent with the message

- wording
- languages
- tongues

A. Outline → telling statistics

3- *Elocutio* (expression)

Figures of	Definition
Thinking	change in words' or propositions' invention and imaginative shape
Meaning (or tropes)	change in words' meaning
Diction	change in words' shape
Elocution	choice of the most suitable or convenient words
Construction	change in words' order inside a sentence
Rhythm	phonic effects

A. Outline → telling statistics

4- *Actio* (execution)

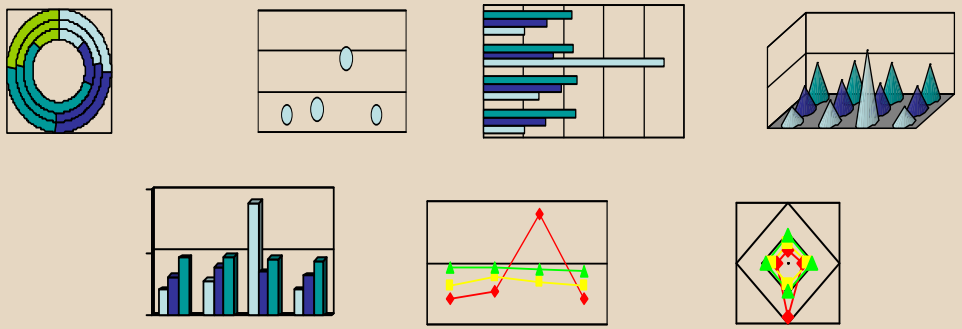
concerns the way in which the telling is managed

in terms of {
- introduction
- developments
- comments
- time space use
- ending

B. Tools → Depicting statistics

Refer to all instruments aimed at depicting statistics

- graphs
- tables
- pictograms



The tools should preserve the message

B. Tools → Depicting statistics

FUNCTIONS

Supporting attention
Activating and building prior knowledge
Minimizing cognitive load
Building mental models
Supporting transfer of learning
Supporting motivation

B. Tools → Depicting statistics

Perception of statistical Graphs

Recognizing the code

Recognizing regularities

Carrying out comparisons and identifying differences

B. Tools → Depicting statistics

Graph Principles

Categories		Principles
Connect with the audience	Message should connect with the goals and interests of your audience.	Relevance
		Appropriate knowledge
Direct and hold attention	Presentation should lead the audience to pay attention to what is important.	Salience
		Discriminability
		Perceptual organization
Promote understanding and memory	Presentation should be easy to follow, digest, and remember.	Compatibility
		Information changes
		Capacity limitations

B. Tools → Depicting statistics

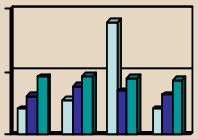
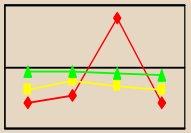
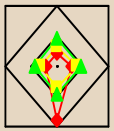
(I) Choosing a graph ...

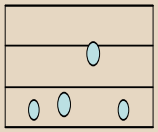
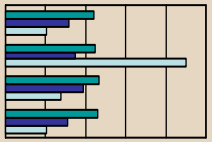
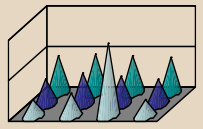
... by taking into account

- **number of involved variables**
- **nature of data (level of measurement)**
- **statistical information to be represented**

... by preferring

- **a simple graph with reference to the audience**
- **a clear graph instead of an attractive one**
- **a correct graph with reference to data**

B. Tools → Depicting statistics

(II) Preparing a graph

Scale definition	**correctly defining and showing scale/s**
Dimensionality	**reducing dimensionality as much as possible by showing few variables for each graph using no meaningless axis**
Colours as statistical codes	**using colours consistently with statistical information**
Rounding off values	**rounding up and down through standard criteria**
Dynamics presentation	**dynamic perspective should reflect a dynamic phenomenon**
Legibility	**few elements as possible. Wise use of legends and captions**

C. Clothes → dressing statistics

Refer to the process of dressing statistics

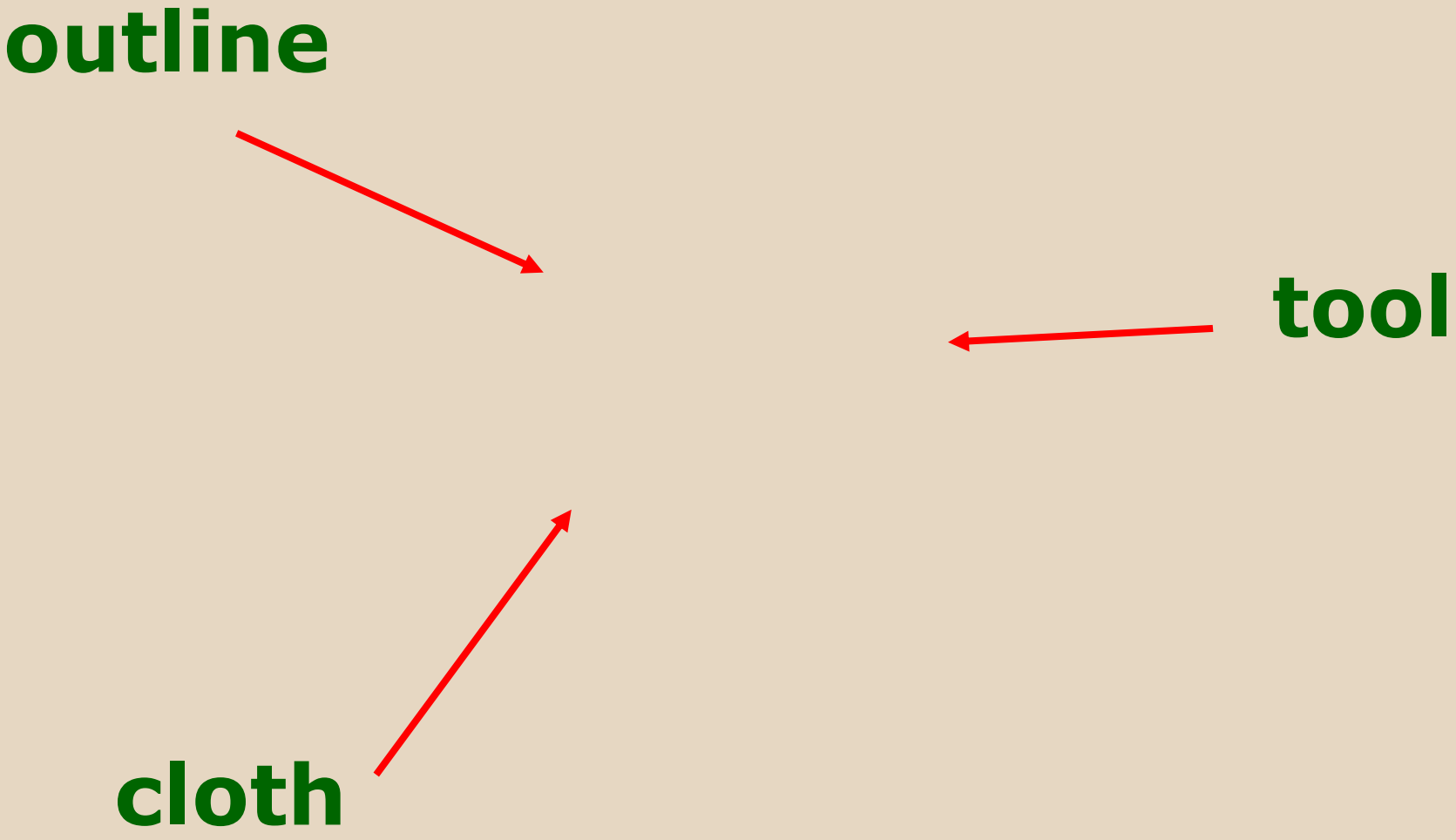
Different aspects:

- text arrangement
- characters and fonts
- colours
- …

With reference to:

- balance
- harmony
- proportion
- elegance
- style

Example



Contents

Contents

3. Assessing the quality of communication in statistics
    1. The conceptual model
    2. The application

1. The conceptual model

A. The dimensions to evaluate
B. The evaluating criteria
C. The components of the transmission process

A. The dimensions to evaluate

1. OUTLINE → telling statistics
2. TOOLS → depicting statistics
3. CLOTHES → dressing statistics

B. The evaluating criteria

They refer to the transmitter's ability to use the codes in terms of

(A) appropriateness → pertinence
(B) correctness → accuracy
(C) clarity

Evaluating scale →

Polarity	Labels	Scores
Bipolar	No	0
	Yes	1

C. The component of the transmission process

(i) **Audience** → tourists, harvesters, miners (*)
(ii) **Channel** → auditory, visual, ….
(iii) **Context** → seminars, conferences, books, booklets, …

But also
(iv) **Topic** }
(v) **Data** } → **message**

C. The component of the transmission process

(i) Audience	- experts - politicians and policy makers - students - statistical data users - not specialized
(ii) Channel	- auditory channel ("listening", requiring oral explanation) - visual channel ("looking", requiring explicative slides) - kinetic channel ("doing", requiring practical exercises)
(iii) Context	- occasions (seminars, conferences, meetings, press conferences, ...) - settings (rooms, tables, ...)

The assessment model

The dimensions of the code	have to be evaluated -through the defined crieria-	with reference to the components of the transmission process
1. Outline 2. Tools 3. Clothes	A. **Appropriateness** (→ pertinence) B. **Correctness** (→ accuracy) C. **Clarity**	i. Audience ii. Channel iii. Context iv. Topic v. Data

2. The application

A. The assessing table
B. Study planning and data collection
C. Data analysis

A. The assessing table

The conceptual model can be consistently assessed by developing an *Assessing Table* through which

each judge can evaluate
presence (1) or absence (0)
….

A. The assessing table

.....

of the criterion

(A) appropriateness (B) correctness (C) clarity

in each code

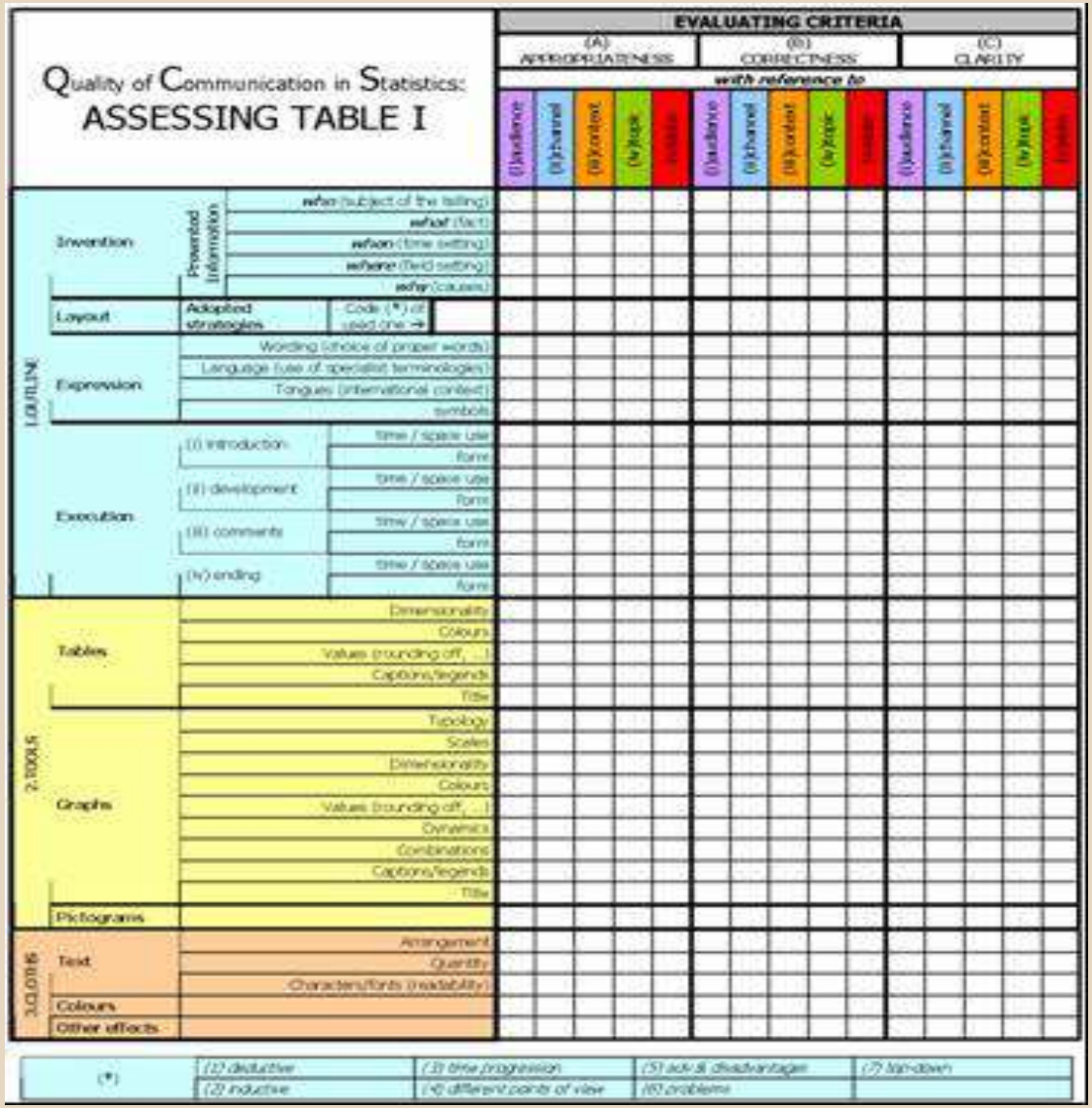
1. outline 2. tools 3. clothes

with reference to

(i) audience (ii) channel (iii) context (iv) topic (v) data

A. The assessing table

Assessing Table I



A. The assessing table

Assessing Table II

synthesis of the previous one

Quality of Communication in Statistics: ASSESSING TABLE II		EVALUATING CRITERIA														
		(A) APPROPRIATENESS					(B) CORRECTNESS					(C) CLARITY				
		with reference to														
		(i)audience	(ii)channel	(iii)context	(iv)topic	(v)data	(i)audience	(ii)channel	(iii)context	(iv)topic	(v)data	(i)audience	(ii)channel	(iii)context	(iv)topic	(v)data
1. OUTLINE	a. Invention															
	b. Layout															
	c. Expression															
	d. Execution															
2. TOOLS	a. Tables															
	b. Graphs															
	c. Pictograms															
3. CLOTHES																

B. The study planning and data collection

Selection of the judges

1. Competence in survey methodology and statistical issues
2. Competence in communication theory

B. The study planning and data collection

Selected publications for the study

(collected at the UNECE Work Session on Communication and Dissemination of Statistics held in Warsaw, Poland – 13-15 May 2009):

- **Central Statistical Office** (2009) *Poland in the European Union*, Central Statistical Office, Warsaw.
- **Eurostat** (2008) *Statistical Portrait of the European Union – European Year of Intercultural Dialogue*, Eurostat, Statistical Books, Luxembourg.
- **Federal Statistical Office** (2009) *Statistical Data on Switzerland*, Federal Statistical Office, NeuChâtel, Switzerland.
- **Kazakhstan Statistics** (2008) *The Statistical Guidebook*, Agency of the Republic of Kazakhstan on Statistics (Astana).
- **ISTAT** (2009) *Italy in Figures*, Rome, Italy
- **United Nations – Economic Commission for Europe** (2009) UNECE. *Countries in Figures*, United Nations, New York – Geneva.

C. Data analysis

OBJECTIVE

assessing each statistical publication
through binary data & ordinal dimensions

PROBLEM

how to combine the evaluations
on each quality dimension
into a final quality assessment

SOLUTION

computing quality assessments
respecting the ordinal nature of data
through a fuzzy approach
based on the use of partial order theory

C. Data analysis

	EVALUATING CRITERIA								
	(A) APPROPRIATENESS			(B) CORRECTNESS			(C) CLARITY		
	with reference to								
	(i)audience	(ii)channel	(iii)context	(i)audience	(ii)channel	(iii)context	(i)audience	(ii)channel	(iii)context
1. OUTLINE	0	1	0						
2. TOOLS	1	1	1						
3. CLOTHES	1	1	1						

Each publication has a sequence of [0/1] for each criterion

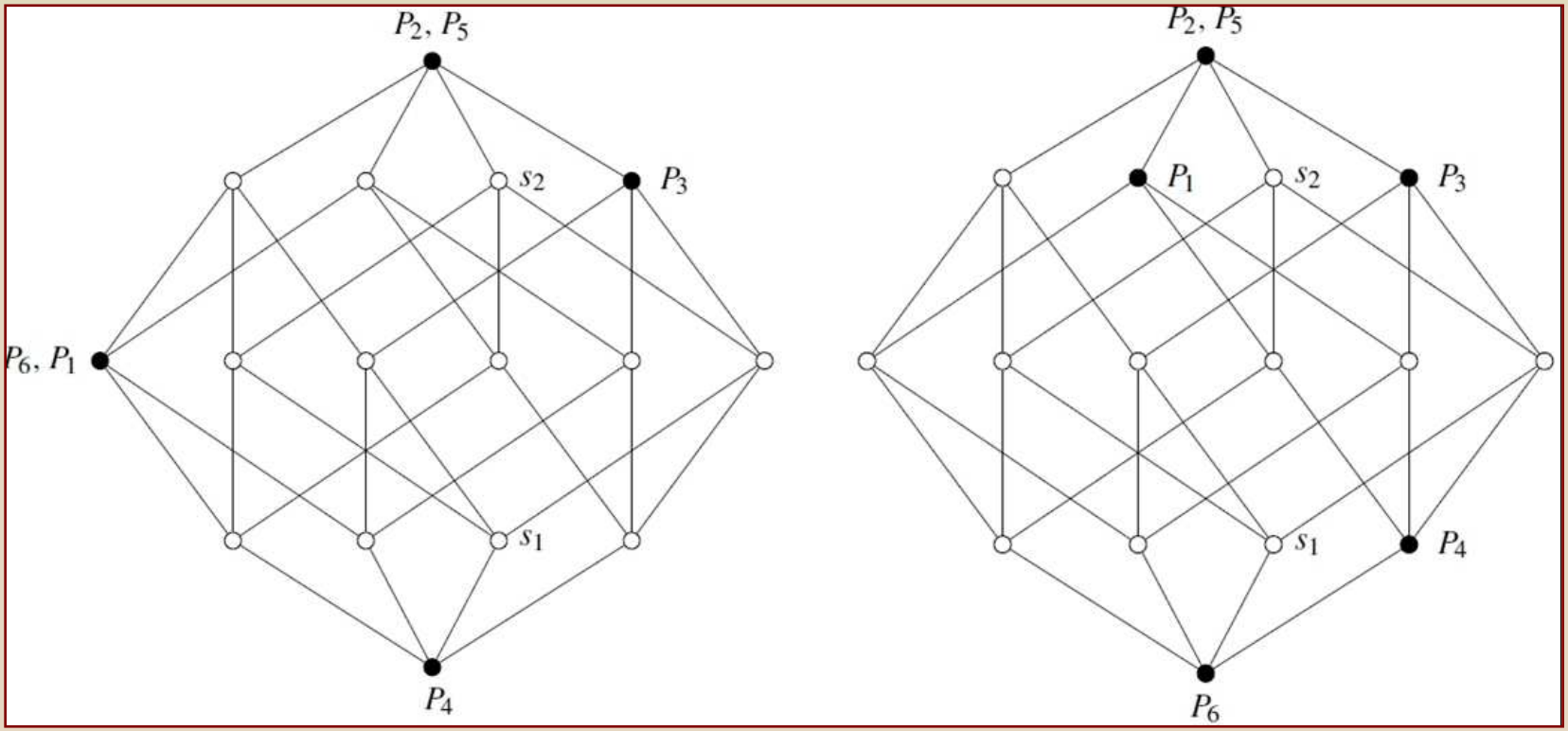
↓

PROFILE

Best configuration → 111111 …
Worst configuration → 000000 …

The analysis was performed for each criterion.
We show just the results concerning
appropriateness *and* clarity.

C. Data analysis



Hasse diagrams of quality configurations

audience appropriateness (left) and audience clarity (right)

for the publication outlines

Linked nodes are ordered from top to bottom.

Not linked nodes represent incomparable quality (appropriateness or clarity) configurations.

C. Data analysis

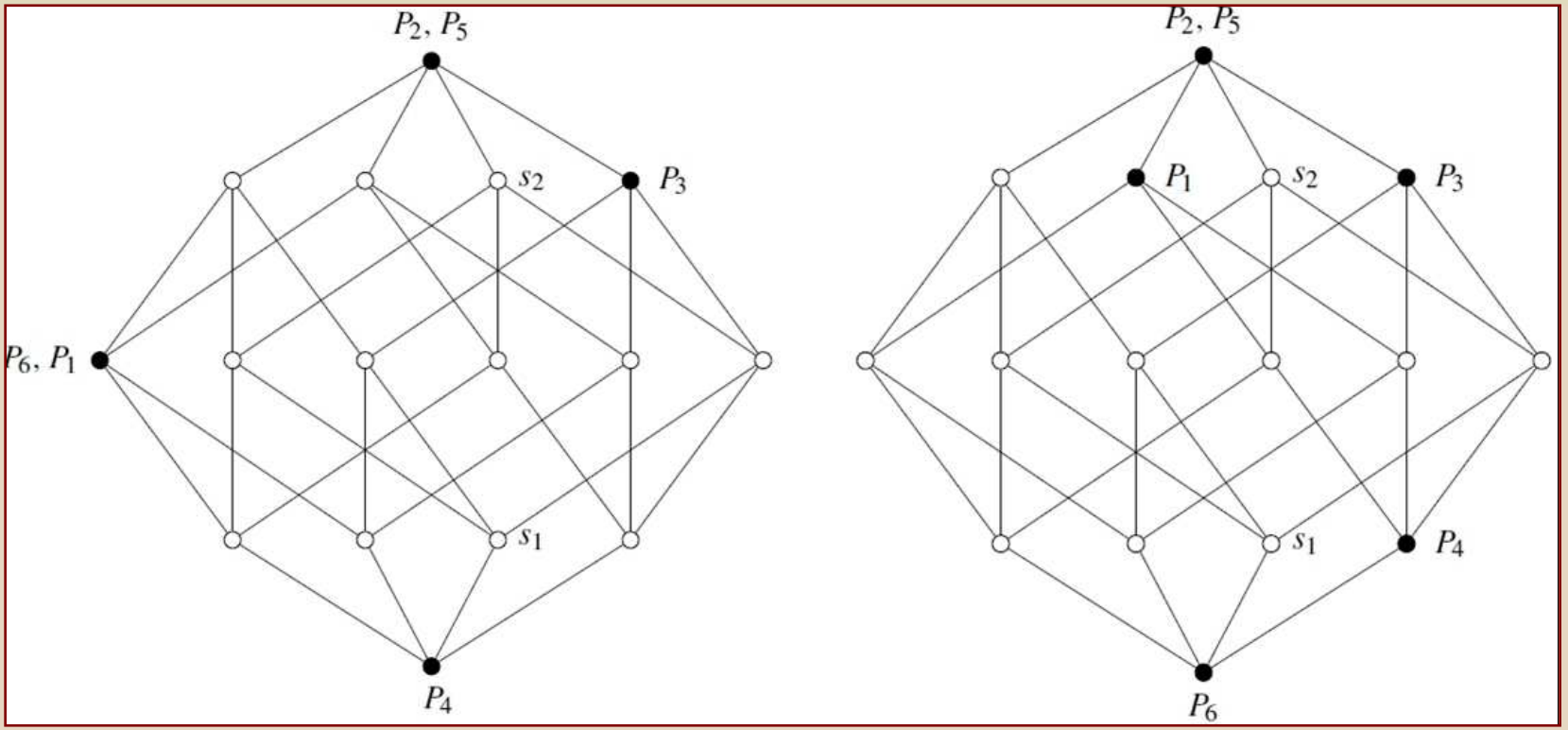
Definition of thresholds (subjective choices)

which element in the sequence is related with

- high quality configuration (quality degree = 1) → $\mathbf{s_2}$
- poor quality configuration (quality degree = 0) → $\mathbf{s_1}$

Given such thresholds, what quality degrees do other configurations receive, in the appropriateness and clarity posets respectively?

C. Data analysis

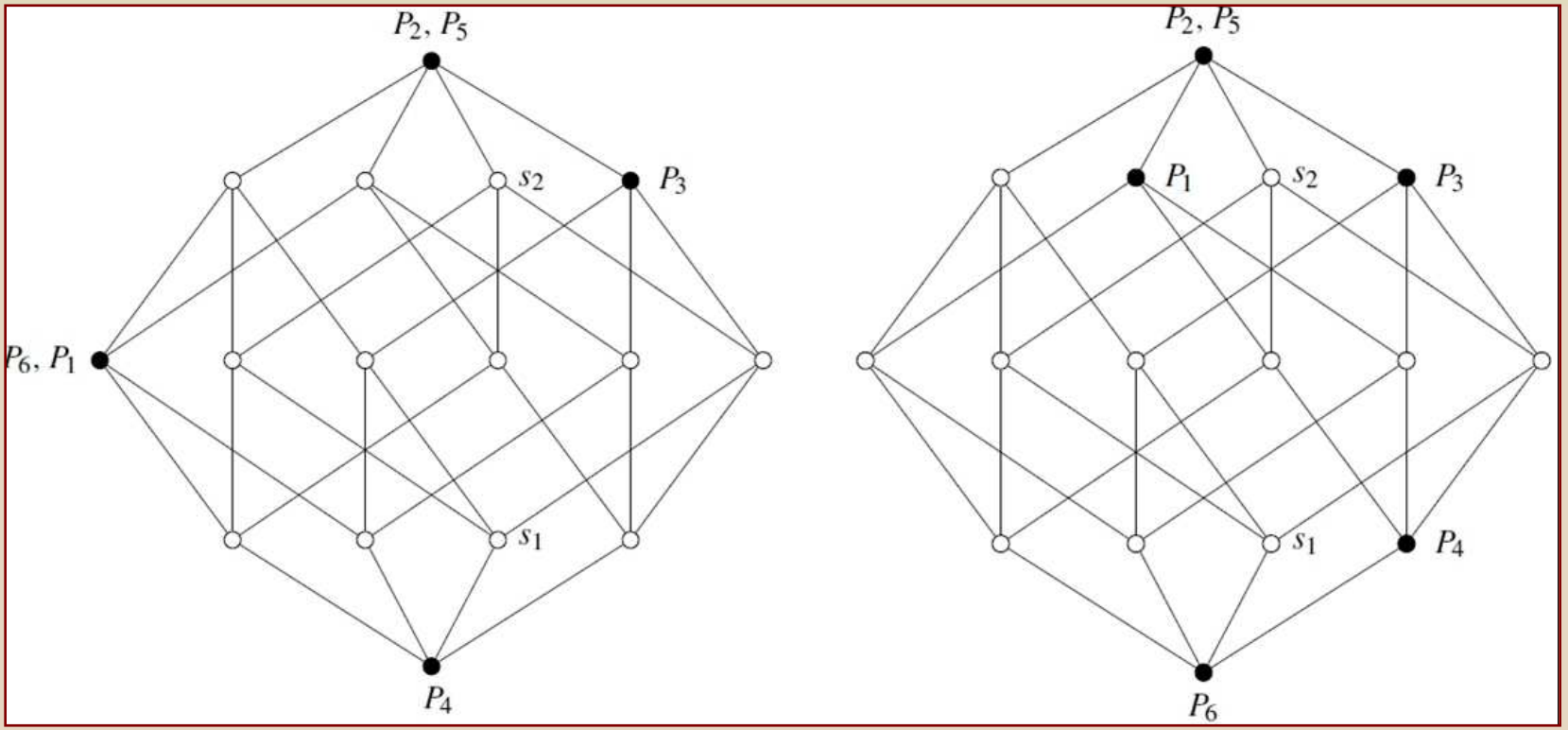


P_2 and **P_5** are above the high quality threshold, in both posets,

↓

they receive quality degree 1 in both appropriateness and clarity

C. Data analysis

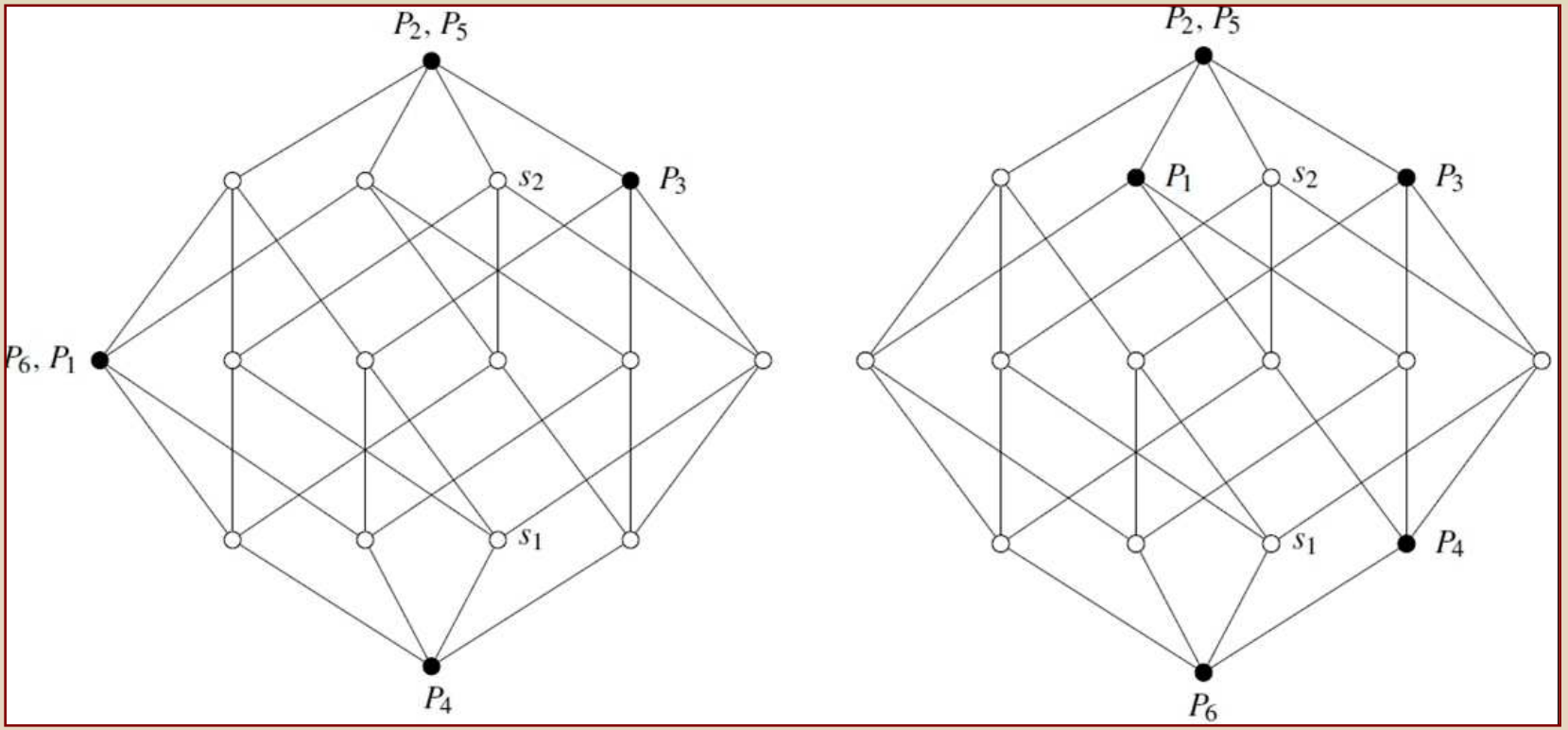


P_4 is below the poor quality threshold, in appropriateness,

⬇

It receives appropriateness degree = 0

C. Data analysis



P_6 is below the poor quality threshold, in clarity,

↓

It receives clarity degree = 0

C. Data analysis

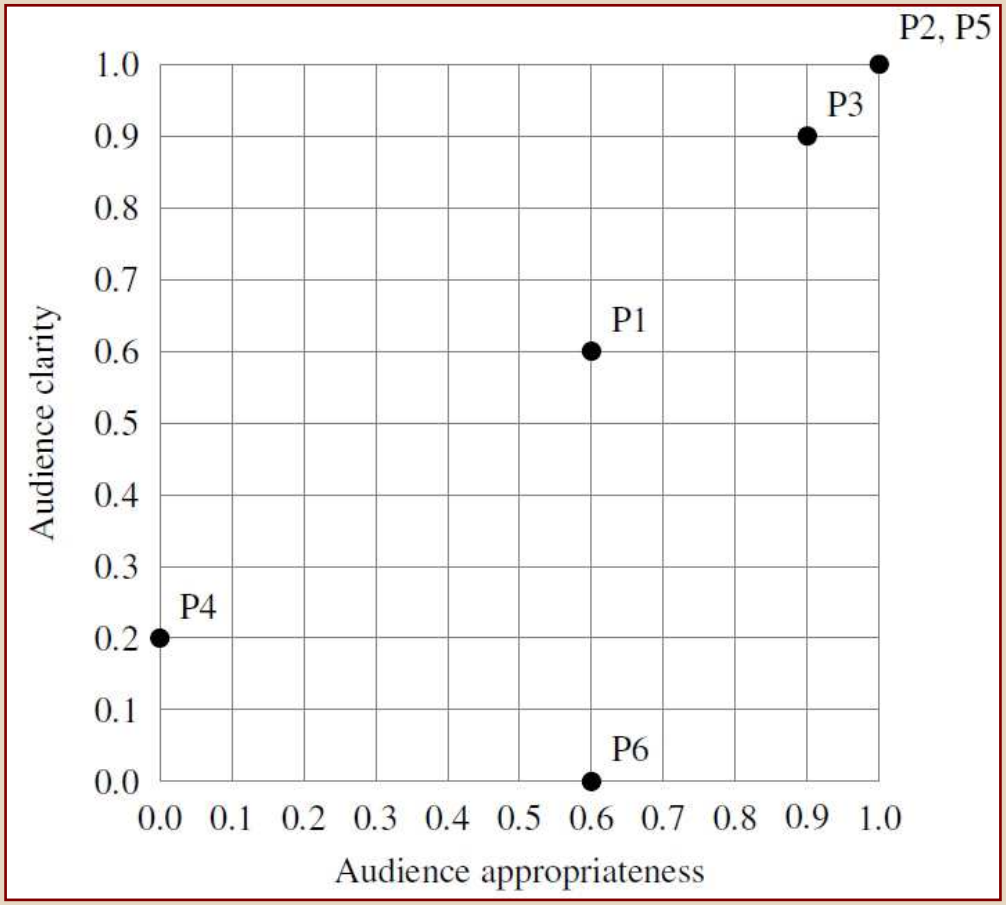
By analysing **how frequently** a configuration is above the **high quality** threshold (or below the poor quality threshold) in the set of complete orders

we can determine **the degree** of **appropriateness** and **clarity** of each configuration (→ publication)

C. Data analysis

Publication	Audience appropriateness	Audience clarity
P1	0.6	0.6
P2	1.0	1.0
P3	0.9	0.9
P4	0.0	0.2
P5	1.0	1.0
P6	0.6	0.0

Final ranking scatterplot



THE WAY FORWARD ...

Goals

- **Improving the assessing model**
- **New applications**
- **Promoting an improvement of statisticians' education by proposing a training module on communication**

The way forward …

Final goal is to assess standardized methods and techniques in order to improve the impact of communication in statistics

Developing and adopting standardized codes allow transmitters to warrant:	
Objectivity	**of data presentation, by avoiding introduction of any subjective component**
Comparability	**between different presentations and along time**
Economicity **and *efficiency***	**in preparing presentation**
Generalization	**by avoiding any kind of "adaptability" of codes to "subjective" messages**
Understanding	**of data structure**

Many thanks for your attention

Filomena Maggino, Marco Fattore, Marco Trapani

Contact: filomena.maggino@unifi.it