Indirect Pairwise Comparison Method

An AHP-based Procedure for Sensory Data Collection and Analysis in Quality and Reliability Applications

FLAVIO S. FOGLIATTO
Federal University of Rio Grande do Sul
Porto Alegre, Brazil

SUSAN L. ALBIN
Rutgers University, USA

Problem Definition

- Sensory evaluation methods:
  - Allow organized collection of information on sensory attributes.
  - Applied in:
    - Product development;
    - On-line, off-line quality control;
    - Product reformulation;
    - Reliability (degradation) analysis.

Problem Definition (*Cont.*)

- Samples in panels are <u>different outcomes of experiments</u>, such as:
  - Product formulations
  - Manufacturing setups
- From panel results we want to build mathematical models.
- Models relate factors varied when preparing samples to corresponding sensory impact.

Indirect Pairwise Comparison Method

- **Key Idea:** use Analytic Hierarchy Process (AHP) principles to guide collection & analysis of sensory data:
  - – AHP mainly used in managerial decision making.
  - – Uses in product development restricted to project selection.

Indirect Pairwise Comparison Method

Main Features

- Sensory panel data elicited from panelists not exposed to intensive training:
  - Result based on empirical evidence.
- Consistency index objectively measures how well panelists perform evaluations:
  - Index allows weighing panelists to favor consistent evaluations.

Related Work

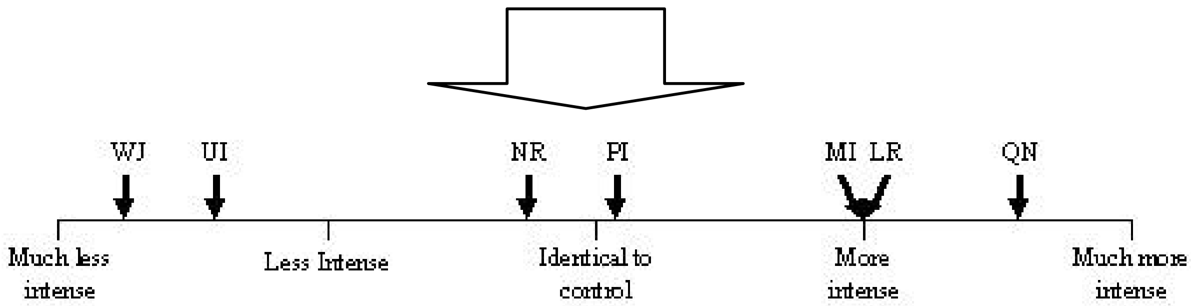
- Fogliatto, Albin & Tepper (*J.Sensory Studies*, 1999) and Fogliatto & Albin (*IIE Transactions*, 2001):
  - First to propose AHP in product development involving sensory variables.
  - Focus was on multivariate optimization problem.

Related Work (*Cont.*)

- Data collection in IPC: similar to Gabrielsen (*FQP*, 2000 & 2001).
- Data analysis in IPC: similar to Qannari *et al*. (FQP, 2000).

IPC Method
Overview

- **Key idea**: quantitatively evaluate intensity of attributes in samples by comparing them to control sample:
  - Panelist presented to entire group of *N* samples.
  - One is identified as *control sample*.
  - Evaluation results recorded on printed scale.



IPC Method
Overview (*Cont.*)

- Then, change control sample and perform evaluations again; each sample will be control at its turn.
- After completing data collection *N* printed scales will be at hand.
- Possible to use procedure using only a subset of samples as controls.

IPC Method - *Overview*
Convert scale marks into numbers

- Arrange results in ($N \times N$) square judgement matrix:
  - Rows labeled 1 to N, each corresponding to a control sample.
  - Entries a_{ij} give result of comparing sample j to control sample i.
  - There will be one matrix per panelist.

IPC Method - *Overview*
Information extracted from judgement matrix

- **Weight vector -** vector of scores corresponding to intensity of each sample:
  - Scores on a continuous scale from 0 to 1 describe attribute intensity or preference.
- **Consistency ratio -** performance measure for each panelist:
  - Describes to what extent transitivity is respected when several samples are evaluated simultaneously by panelist.

IPC Method - *Overview*
Advantages

- Yields continuous data usable for model building purposes.
- Allows measuring of panelists' efficiency through consistency ratios:
  - Panel leader able to assess effectiveness of training practices & combine evaluations from panelists using consistency ratios as weights.
- Less inconsistency, since all samples available at once for comparison.

IPC Method - *Overview*
Major Drawback

- Panelists fatigue from simultaneous presentation of samples:
  - that leads to more evaluation sessions and higher data collection costs.
  - Limitation present in any sensory data collection method; gain in consistency provided by IPC may outdo this limitation

Steps for Data Collection

- Organize samples in a judgement matrix:

Samples

	1	2	- - -	N
1	1	a_{12}	- - -	a_{1N}
2	a_{21}	1	- - -	a_{2N}
⋮	⋮	⋮	⋱	⋮
N	a_{N1}	a_{N2}	- - -	1

  – Each matrix row constitutes a separate test: all samples compared to the sample corresponding to row label i, $i = 1,\ldots, N$.

  – Panelist performs $N - 1$ comparisons to complete single test.

  – In a complete run of IPC N tests will be performed.

A simple way to instruct panelists

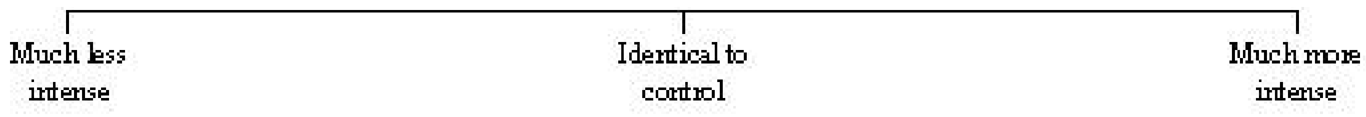
- "You will be presented with a group of samples, one of them identified as the control sample. Your task is to tell how intense they appear to you in comparison with the control. The intensity of the control sample corresponds to the center of the scale in front of you. Samples that are more intense than the control must have their codes marked on the right-hand side of the scale accordingly, and those less intense than the control on the left-hand side. When two samples appear to be equally intense, write their codes at the same spot on the scale."

Evaluations marked on printed scale

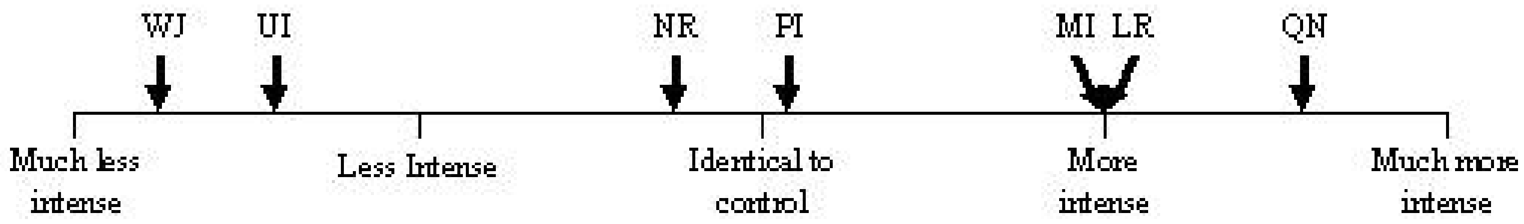
- We suggest using a 15 cm line, w/ 3 marks; in case of many samples, use larger scale.
- Use printed marks as anchors rather than numbers.
- Evaluations later converted into numbers ranging from $^1/_9$ to 9, which are written in rows of judgement matrix.

Printed Scale

- Proposed scale:



- Example (w/ 5 anchors and 8 samples):



In a complete run of procedure each panelist performs *N* tests

- One printed scale w/ results from each test.
- Converting scale marks into [$^1/_9$, 9] values:

$$a_{ij} = \begin{cases} 1 + (8z/s) & z \geq 0 \\ [1 - (8z/s)]^{-1} & z < 0 \end{cases}$$

Half-length of the scale.

Location of mark on scale, in *cm*; sign of *z* is negative to the left of scale center and positive to the right.

Result of comparing sample *j* against control sample *i*.

S_4: $^1/_5$ S_3: $^1/_3$ S_2: 3

$^1/_9$ 1 9

Suppose sample #1 is control

Samples	1	2	3	4
1	1	3	1/3	1/5

Indirect Pairwise Comparison Method

Fogliatto & Albin

UFRGS

Indirect Pairwise Comparison Method

Fogliatto & Albin

Matrix will contain two comparisons for samples *i* and *j*

- Comparisons are a_{ij} and $1/a_{ji}$; they are likely to be non-identical.
- Data analysis to follow requires judgement matrices to be reciprocal; to overcome that, calculate midpoint values of a_{ij} and $1/a_{ji}$, and let a_{ij} be the resulting value.

Samples	1	2	3	4
1	1	3	1/3	1/5
2	1/2.5	1	2	1/1.5

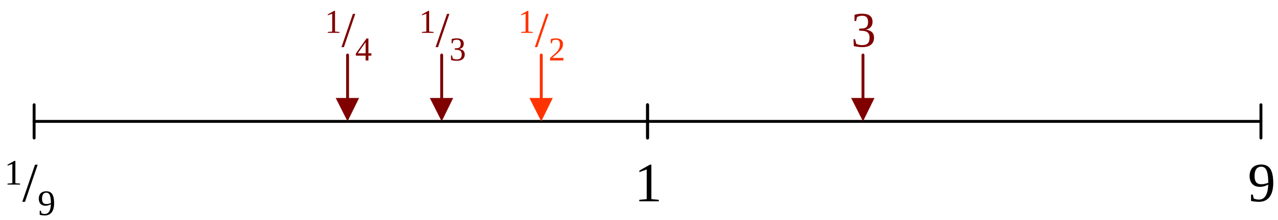
- Final corrected judgement matrix will be obtained s.t. element a_{ij} is reciprocal of a_{ji}.

How to calculate midpoints

- Convert a_{ij}´s back to z's.
- Find average of resulting z's and convert it back to [$^1/_9$, 9]-scale to find midpoint.
- Calculation works for any number of a_{ij}´s.

Example
Determining midpoint of 3 outcomes

- Values are 3, $^1/_3$ and $^1/_4$ (under perfect consistency, evaluation outcomes should be identical).
- Converting back to *z* yields 2.5, –2.5, and –3.75; the average is –1.25.
- The resulting midpoint is $^1/_2$.



Indirect Pairwise Comparison Method

Fogliatto & Albin

Data collection yields midpoint matrices

- $\bar{\mathbf{A}}$ = midpoint matrix:
  - Diagonals are 1.
  - Elements above diagonals are the midpoints.
  - Elements below diagonals are reciprocal of elements above diagonal.
- Example (N=3):

$$\bar{\mathbf{A}} = \begin{bmatrix} 1 & \bar{a}_{12} & \bar{a}_{13} \\ 1/\bar{a}_{12} & 1 & \bar{a}_{23} \\ 1/\bar{a}_{13} & 1/\bar{a}_{23} & 1 \end{bmatrix}$$

How many samples should we presented in a test?

- In average less than 10, but that depends on sense being stimulated.
- Guide table was gathered from magnitude estimation literature (mostly):

Sense Stimulated	*Absolute Evaluations Number of Samples*	*Relative Evaluations Number of Samples*
Taste	4 – 6	6 – 8
Sight	7 – 9	6 – 10
Smell	-	8 – 13
Hearing	4 – 8	6 – 10

Adapting procedure to reduce number of tests

- From a given judgement matrix row all others may be derived, using expressions:

$$a_{ij} = 1/a_{ji}$$
$$a_{ij} = a_{ik} \times a_{kj}$$

- Ex.: only row #1 is given. Element a_{34} in row #3 is given by $a_{34} = a_{31} \times a_{14} = 1/a_{13} \times a_{14}$.
- Resulting matrix is perfectly consistent.

Pick some samples and use as control samples

- Choose *M* out of *N* possible tests.
- Write results from tests in *M* separate matrices:
  - Each will have a single row w/ numbers.
  - Complete entries in matrices using expressions.
- If panelist is perfectly consistent, matrices will be identical; otherwise, determine matrix of midpoints.

Data Analysis

- Weight vector: vector of intensity weights, each corresponding to a sample.
- Consistency ratio: measures transitivity.
  - If A > B and B > C, then transitivity is respected if A > C.

Weight vector, w

- Given by matrix's principal eigenvector (Saaty, 1977):

$$\bar{\mathbf{A}}\mathbf{w} = l_{\max}\mathbf{w}$$

$l_{\max}$ → Matrix's largest eigenvalue

Panelist A	*1*	*2*	*3*
Sample 1	1	4	1
Sample 2	1/4	1	1/9
Sample 3	1	9	1

⇒

Panelist A	**w**
Sample 1	0.29
Sample 2	0.23
Sample 3	0.48

Why does vector of weights = Eigenvector of l_{max}?

- For each entry (i, j) of $\bar{\mathbf{A}}$:

$$\bar{a}_{ij} = w_i / w_j \quad \rightarrow \quad \bar{a}_{ij}(w_i / w_j) = 1$$

- Summing over any given row i of $\bar{\mathbf{A}}$:

$$\sum_{j=1}^{N} \bar{a}_{ij} w_j \frac{1}{w_i} = N \ , \ i = 1 \mathbf{K} \, N$$

or

$$\sum_{j=1}^{N} \bar{a}_{ij} w_j = N w_i \ , \ i = 1 \mathbf{K} \, N$$

Why does vector of weights = Eigenvector of l_{max}?

- For all rows of $\bar{\mathbf{A}}$:

$$\bar{a}_{11}w_1 + \bar{a}_{12}w_2 + \mathbf{K} + \bar{a}_{1n}w_n = Nw_1$$

$$\bar{a}_{21}w_1 + \bar{a}_{22}w_2 + \mathbf{K} + \bar{a}_{2n}w_n = Nw_2$$

$$\mathbf{M} \qquad\qquad \mathbf{M}$$

$$\bar{a}_{n1}w_1 + \bar{a}_{n2}w_2 + \mathbf{K} + \bar{a}_{nn}w_n = Nw_n$$

$$\bar{\mathbf{A}}.\mathbf{w} = N.\mathbf{w}$$

When matrix is perfectly consistent, $\lambda_{max} = N$ as seen next.

Relationships between elements in a consistent matrix

- If $\bar{\mathbf{A}}$ is consistent, then:
  - Elements *i*, *j*, and *k* relate such that:

$$\bar{a}_{ij} = \bar{a}_{ik} \times \bar{a}_{lk}$$

  - From any row of matrix, all other rows can be derived.
  - All but one of $\bar{\mathbf{A}}$'s eigenvalues (λ) are zero:
    - $\lambda_{max} = N$
    - all other $\lambda = 0$.

$$Trace(\bar{\mathbf{A}}) = \sum_{i=1}^{N} l_i.$$

Consistency measure for $\bar{\mathbf{A}}_{NxN}$

- If $\bar{\mathbf{A}}$ consistent, $\lambda_{max} = N$ and all others eigenvalues are 0.
- In general, $\lambda_{max} \geq N$.
- Consistency Index, *CI*:

$$CI = (\, l_{max} - N\,)/(N - 1)$$

Consistency Ratio and Random Index

- Consistency ratio: $CR = \frac{CI}{RI}$ ← Random index

- How to obtain RI?

  For N = 3,4,5,6,... generate 500 random (reciprocal) matrices.

  ❷ Determine their CI's.

  Calculate average CI → Random Index.

Matrix consistent if CR £ 0.1.

Indirect Pairwise Comparison Method

Fogliatto & Albin

Case Example 1: Quality application

Finding the best formula for a pet food

- ***Objective***: find alternative lower cost formula for dog biscuit.
- Samples obtained from mixture design w/ 10 runs:
  - % of 3 ingredients and biscuit thickness (process variable) are the control variables.
  - Mixture variables are dependent: analysis performed after transformation.

Design matrix and responses

Run	1	2	3	4	5	6	7	8	9	10
W1	-1	-1	1	1	0	-1.41	1.41	0	0	0
W2	-1	1	-1	1	0	0	0	-1.41	1.41	0
Thick	-1	-1	-1	-1	-1	1	1	1	1	1
I1	46.4	32.8	16.6	39.7	23.6	7.4	30.5	14.3	0.7	23.6
I2	16.6	32.8	46.4	7.4	23.6	39.7	0.7	14.3	30.5	23.6
I3	7.3	4.7	7.3	23.2	23.2	23.2	39.1	41.7	39.1	23.2

- Two panel responses considered here:
  - Texture (baked dough consistency and crackiness).
  - General appearance (biscuit color, integrity, etc.)

Five panelists evaluated samples

- Panelists familiar w/ product and desired sensory characteristics:
  - samples compared w/ control sample regarding compliance to target characteristics.
- Three samples used as control samples:
  - 3 judgement matrices obtained for each panelist.
  - 3 replications available on each pair of samples.

Data collected: *some examples*

Upper diagonals are midpoints of 3 matrices from panelist 1

1st matrix obtained from panelist 1

		1	2	3	4	5	6	7	8	9	10
	1	1.00	8.03	1.08	0.97	1.05	0.93	0.96	4.21	0.90	67.21
	2	0.12	1.00	0.14	0.12	0.13	0.12	0.12	0.53	0.11	8.37
	3	0.92	7.40	1.00	0.90	0.97	0.86	0.88	3.89	0.83	61.97
	4	1.03	8.23	1.11	1.00	1.08	0.95	0.98	4.32	0.92	68.96
$A_{11}^{(1)}$	5	0.95	7.61	1.03	0.92	1.00	0.88	0.91	4.00	0.85	63.71
	6	1.08	8.65	1.17	1.05	1.14	1.00	1.03	4.54	0.97	72.45
	7	1.04	8.37	1.13	1.02	1.10	0.97	1.00	4.40	0.94	70.12
	8	0.24	1.90	0.26	0.23	0.25	0.22	0.23	1.00	0.21	15.95
	9	1.11	8.93	1.21	1.08	1.17	1.03	1.07	4.69	1.00	74.78
	10	0.01	0.12	0.02	0.01	0.02	0.01	0.01	0.06	0.01	1.00

Midpoint judgement matrix for panelist 1

6	3.22	6.86	10.86	10.33	7.08	1.00	0.84	2.69	1.30	31.01
7	7.71	13.61	20.91	4.94	6.99	1.19	1.00	3.56	1.20	28.58
8	0.81	2.66	5.25	4.55	3.38	0.37	0.28	1.00	0.30	9.52
9	7.27	13.20	20.10	3.33	6.15	0.77	0.84	3.23	1.00	29.17
10	0.04	0.12	0.05	0.04	0.05	0.03	0.04	0.11	0.03	1.00

Two other matrices were obtained from tests were samples #4 and #10 were controls.

Lower diagonals are the reciprocals of upper diagonals

Same type of data obtained from other 4 panelists.

Summary of results
Weight vectors and CRs

- **Texture:** best samples are #7 & 9; best panelist is #5.

	1	2	3	4	5	6	7	8	9	10	CR
w_{11}	.074	.029	.014	.059	.025	.22	.267	.078	.230	.004	.133
w_{21}	.036	.045	.044	.058	.018	.084	.136	.155	.396	.028	.020
w_{31}	.061	.079	.065	.110	.045	.149	.139	.102	.167	.082	.005
w_{41}	.052	.048	.041	.113	.069	.022	.182	.145	.279	.048	.008
w_{51}	.100	.077	.088	.065	.059	.117	.133	.140	.186	.036	.001

- **Appearance:** best samples are #3 & 7; best panelist is #5.

	1	2	3	4	5	6	7	8	9	10	CR
w_{12}	.070	.010	.342	.036	.322	.049	.081	.064	.023	.003	.082
w_{22}	.145	.043	.166	.087	.076	.106	.094	.158	.112	.012	.013
w_{32}	.046	.059	.049	.116	.120	.077	.298	.094	.129	.013	.021
w_{42}	.039	.041	.049	.146	.054	.147	.179	.151	.178	.017	.002
w_{52}	.103	.083	.172	.083	.103	.079	.145	.117	.101	.015	.001

Summary of results
Regression weights and final vectors

- Weights:
  - **Excellence** was a subjective evaluation on panelists given by panel leader (30%).
  - **Reciprocals of CRs** also used (70%).

k	Exc	$1/CR_1$	$1/CR_2$
1	0.04	0.01	0.01
2	0.12	0.05	0.04
3	0.20	0.17	0.02
4	0.28	0.11	0.25
5	0.36	0.67	0.68

- Final weight vectors (combining over all panelists):

$$\mathbf{w}_1 = [0.080;\ 0.069;\ 0.072;\ 0.080;\ 0.055;\ 0.101;\ 0.145;\ 0.134;\ 0.212;\ 0.045]$$

$$\mathbf{w}_2 = [0.084;\ 0.066;\ 0.133;\ 0.102;\ 0.093;\ 0.098;\ 0.161;\ 0.126;\ 0.123;\ 0.015]$$

Indirect Pairwise Comparison Method

Fogliatto & Albin

Summary of results
Regression models and new formulation

- Regression Models:

$$Texture = 0.072 + 0.0287\,Thick + 0.035(W_2 \times W_2) \qquad (R^2 = 0.64)$$

$$Appearance = 0.054 + 0.0217W_1 + 0.0299(W_1 \times W_2) + 0.0274(W_2 \times W_2) \qquad (R^2 = 0.62)$$

- Best overall product:
  - $I_1 = 3.96\%$; $I_2 = 9.43\%$; $I_3 = 56.61\%$
  - $Thickness = +1$
- New product formulation yielded a 18% cost reduction.

UFRGS

Indirect Pairwise Comparison Method

Fogliatto & Albin

Case Example 2: Reliability application
Overview

- Warranty modeling of plastic tiles used in swine maternity flooring. Warranty claims led to a new recipe for tiles.
- *Our task*: determine proper warranty length for the new product.
- 5 panelists performed degradation evaluations on samples from accelerated tests comparing against threshold control sample.

Conclusion
Future work

- Midpoint matrices do not use variance information in the sample. Use rank reversal probabilities to take that into account.
- RIs obtained from simulations using all samples as controls. Run simulations to find RIs for cases where $M < N$.