

# Improving subjective estimations using the paired comparisons method



# ERICSSON

International Forum on COCOMO  
and  
Software Cost Estimation

Eduardo Miranda  
October 24-27, 2000

# Agenda

- Subjective estimates
- Estimating person-hours
- Can we do better? - The paired comparison method
- The process
- Example
- Evaluation
- Why does it work?
- Summary & further research

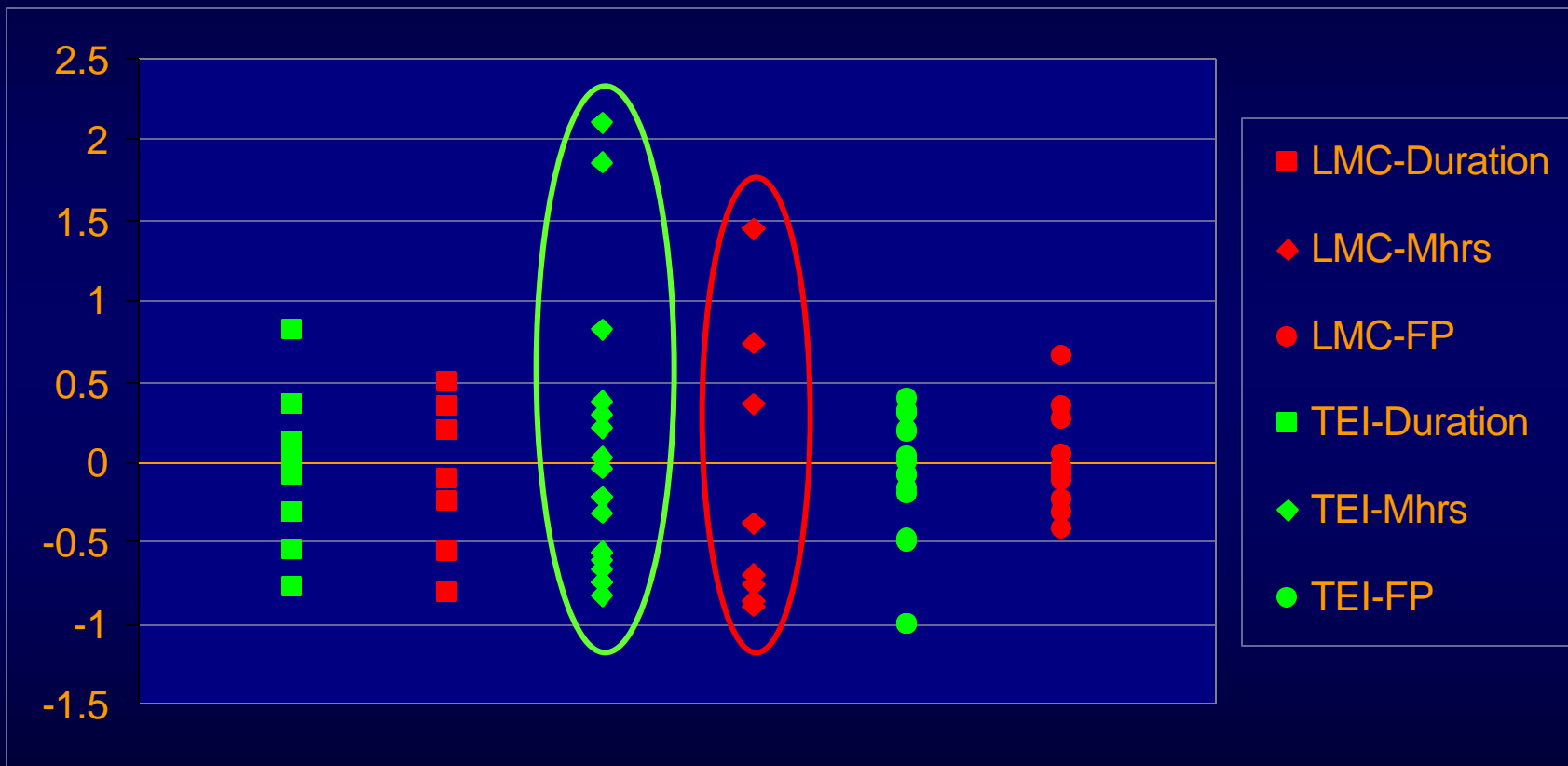


# Subjective estimations are a way of life for many activities and organizations

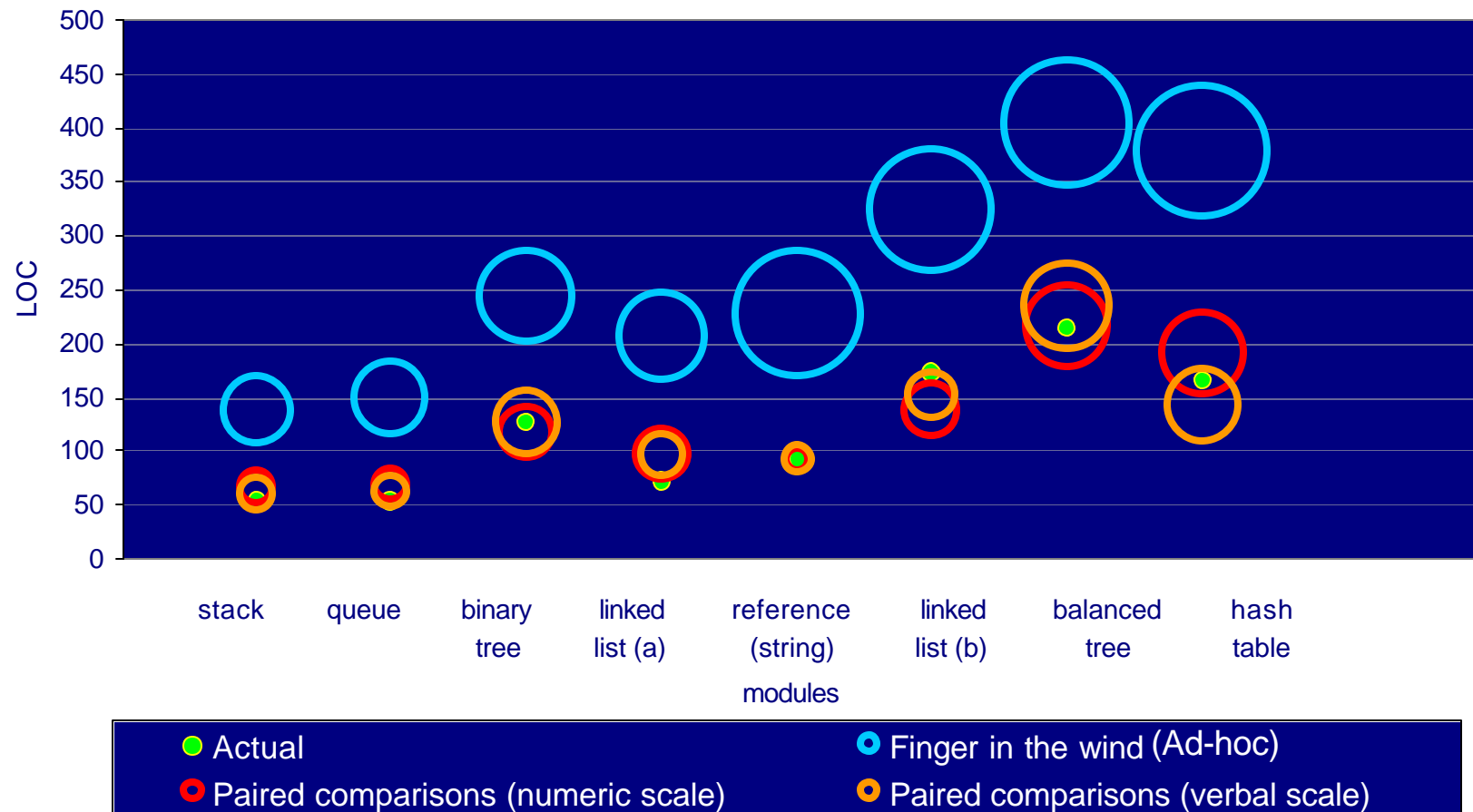
- Lack of necessary information at the beginning of the project
- Vocabulary foreign to the stakeholders
- Some techniques are domain specific
- Require significant time and effort



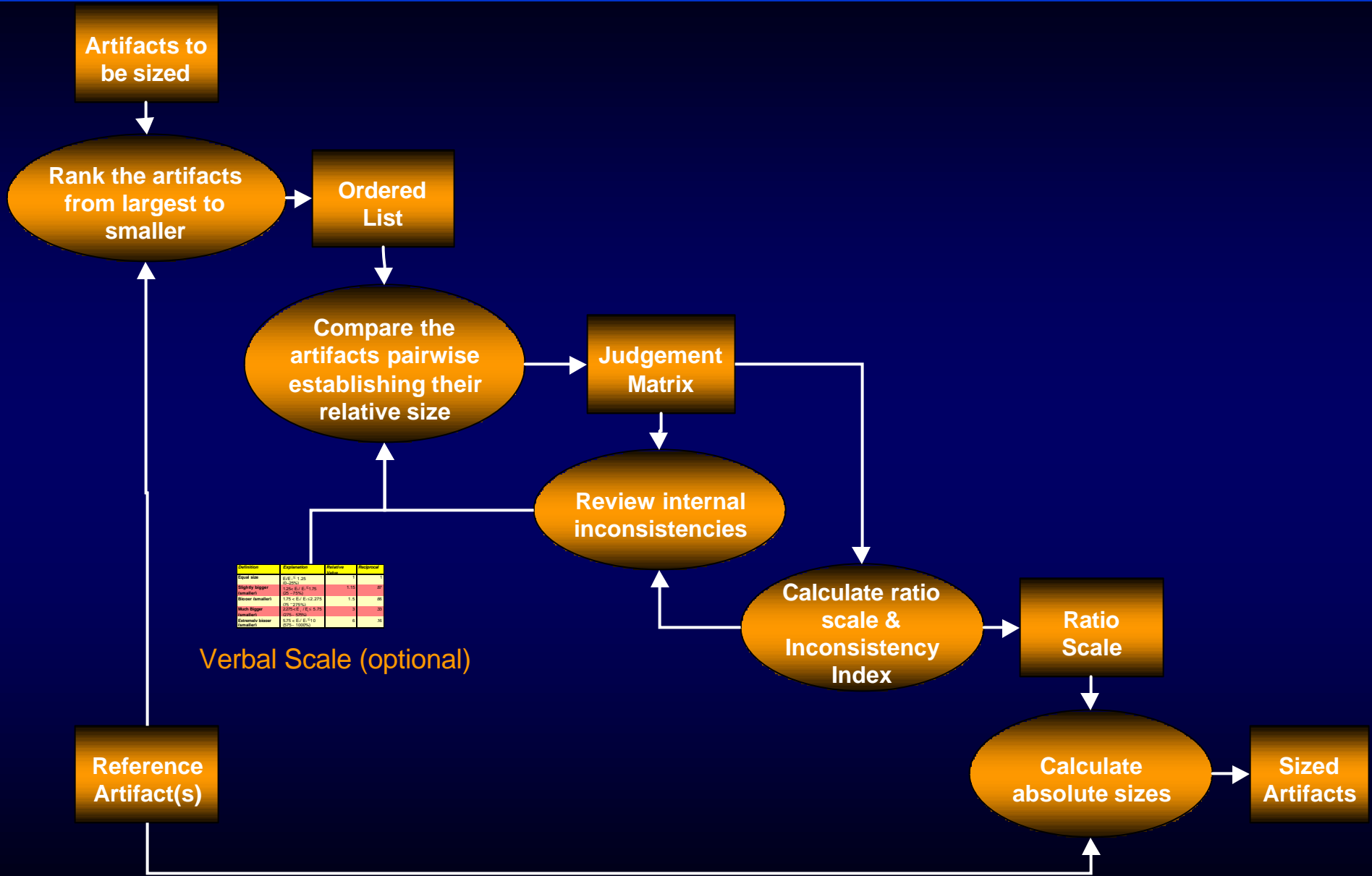
# Estimating person-hours



# Can we do better? - The paired comparison method



# The paired comparisons process



# The math - judgement matrix

Artifacts	g	a	b	c	d	l	j	h	e	f
g		2	2.2	3	3	4	4.5	6	7	9
a			1	1	2.2	3	3	3	5	6
b				1.1	1.2	1.3	5	6	6	7
c					2	2	2	2.3	4.5	5
d						1	1.5	2	3.4	4
l							3	3	3.3	4
j								4	4	4.5
h									1	2
e										2
f										

$$A^{n \times n} = [a_{ij}] = \begin{cases} a_{ij} = \frac{s_i}{s_j} & , \text{How much bigger (smaller) is } E_i \text{ with respect to } E_j \\ a_{ii} = 1 & , \text{Every element has the same size as itself} \\ a_{ji} = \frac{1}{a_{ij}} & , \text{If } E_i \text{ is } a_{ij} \text{ times bigger (smaller) than } E_j, \text{ then } E_j \text{ is } 1/a_{ij} \text{ times smaller} \\ & \text{(bigger) than } E_i \end{cases}$$

## The math - ratio and size calculations

$$(a) \quad n_i = \sqrt[n]{\prod_{j=1}^n a_{ij}}$$

$$(b) \quad r_i = \frac{n_i}{\sum_{l=1}^n n_l}$$

$$(c) \quad \begin{bmatrix} s_k * \frac{r_1}{r_k} \\ s_k * \frac{r_2}{r_k} \\ \dots \\ s_k \\ \dots \\ s_k * \frac{r_n}{r_k} \end{bmatrix}$$

$$(d) \quad InconsistencyIndex = \frac{\sqrt{\sum_{i=1}^n \sum_{j>i}^n \left( \ln a_{ij} - \ln \frac{n_i}{n_j} \right)^2}}{(n-1)(n-2)} \cdot 2$$

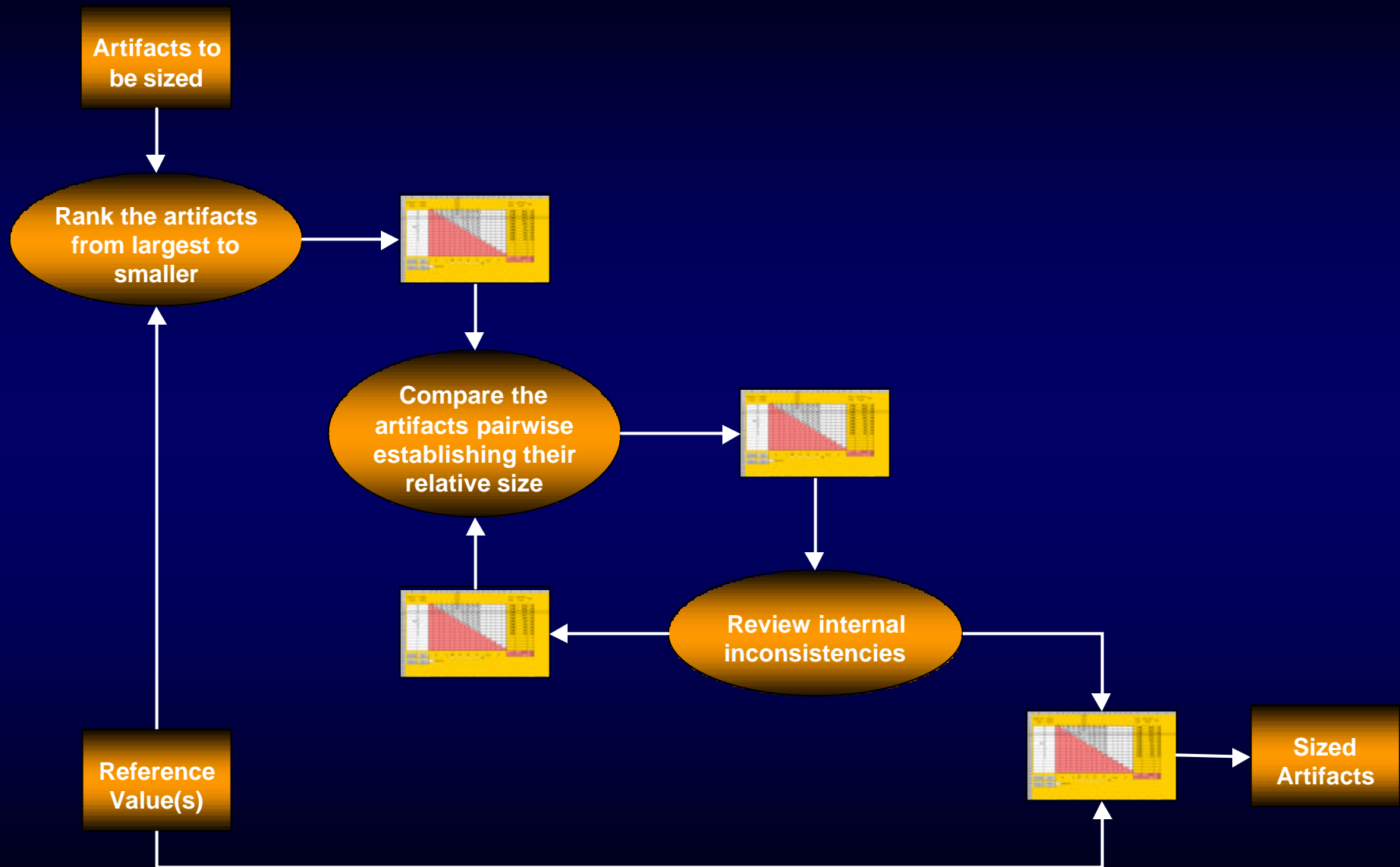
Crawford and Williams' procedure. (a) Row's geometric mean, (b) Ratio scale, (c) Size vector, (d) Inconsistency Index



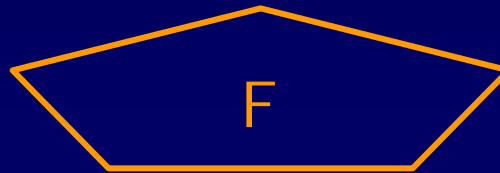
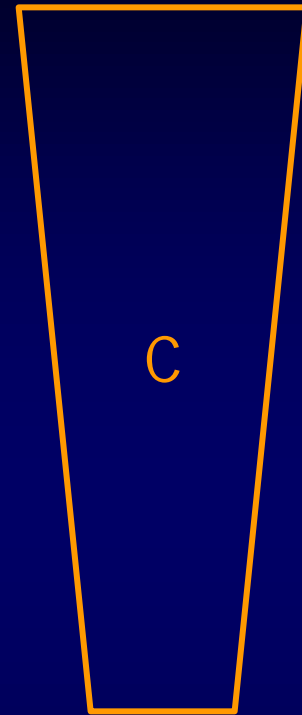
## Verbal scales

<b>Definition</b>	<b>Software Relative Value</b>	<b>AHP Relative Values</b>
Equal size	1	1
Slightly bigger (smaller)	1.25	3
Bigger (smaller)	1.75	5
Much Bigger (smaller)	4	7
Extremely bigger (smaller)	7.5	9

# MinimumTime - automating the process



# Example



# Evaluation

- Stack
- Queue
- Binary Tree
- Hash Table
- Balanced Tree
- Linked List (a)
- Linked List (b)
- String Manipulation

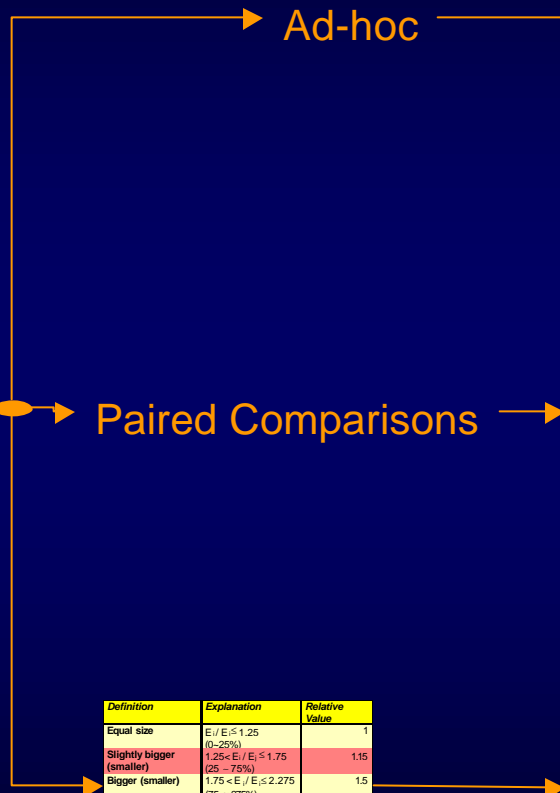
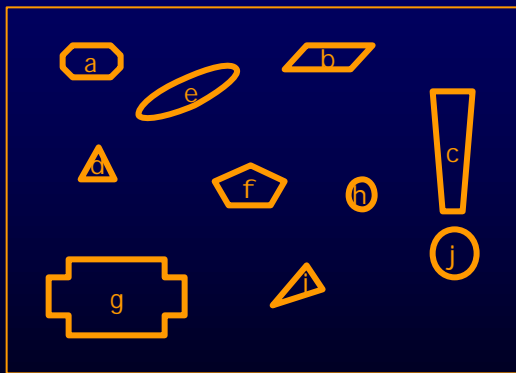


Figure	Size
a	
b	
c	
d	
j	
h	
l	

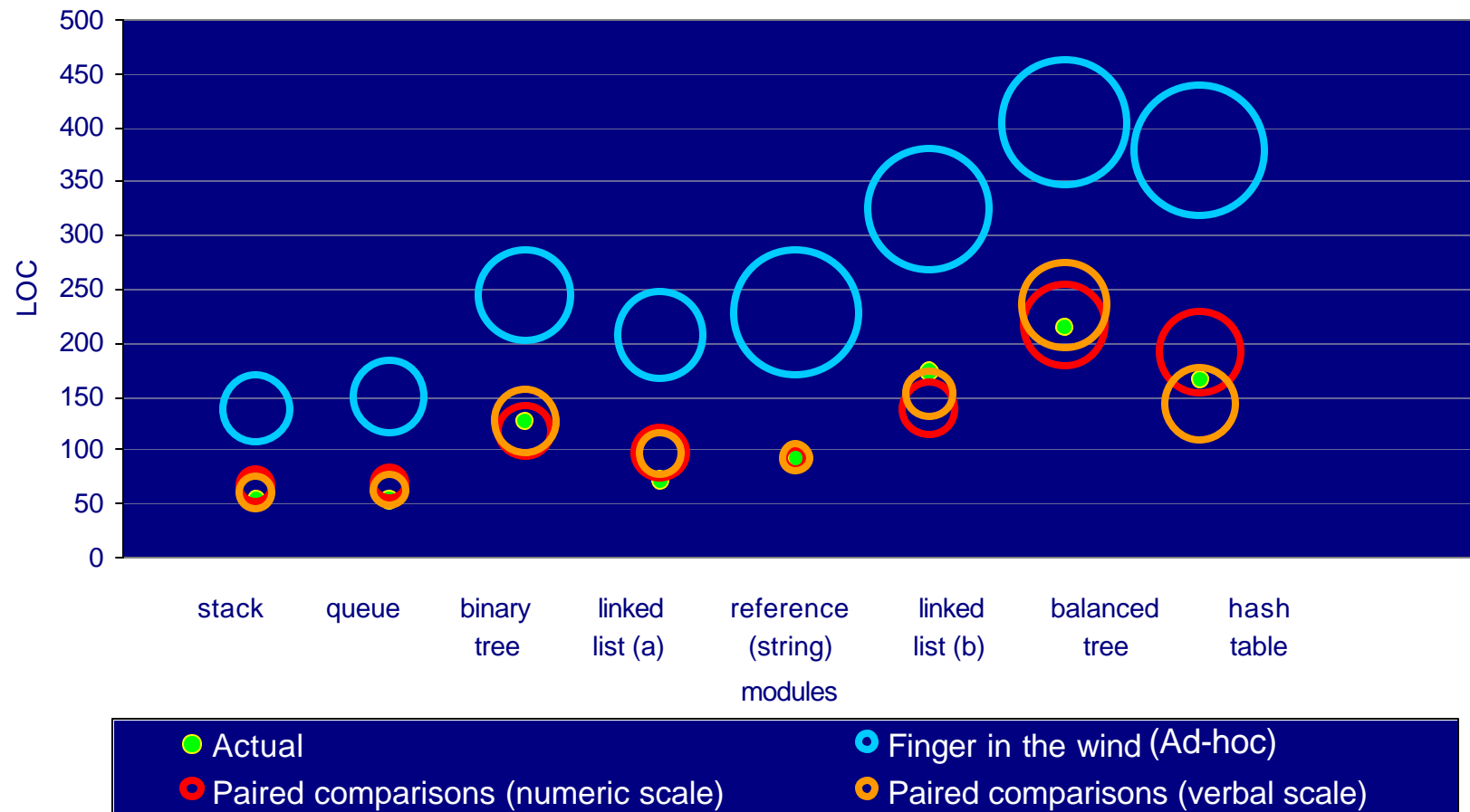
Figure	a	g	f	c	j	b
d						
a						
g						
f						
c						
j						
b						

Definition	Explanation	Relative Value
Equal size	$E_i / E_j \leq 1.25$ (0-25%)	1
Slightly bigger (smaller)	$1.25 < E_i / E_j \leq 1.75$ (25-75%)	1.15
Bigger (smaller)	$1.75 < E_i / E_j \leq 2.275$ (75-275%)	1.5
Much Bigger (smaller)	$2.275 < E_i / E_j \leq 5.75$ (275-575%)	3
Extremely bigger (smaller)	$5.75 < E_i / E_j \leq 10$ (575-1000%)	6

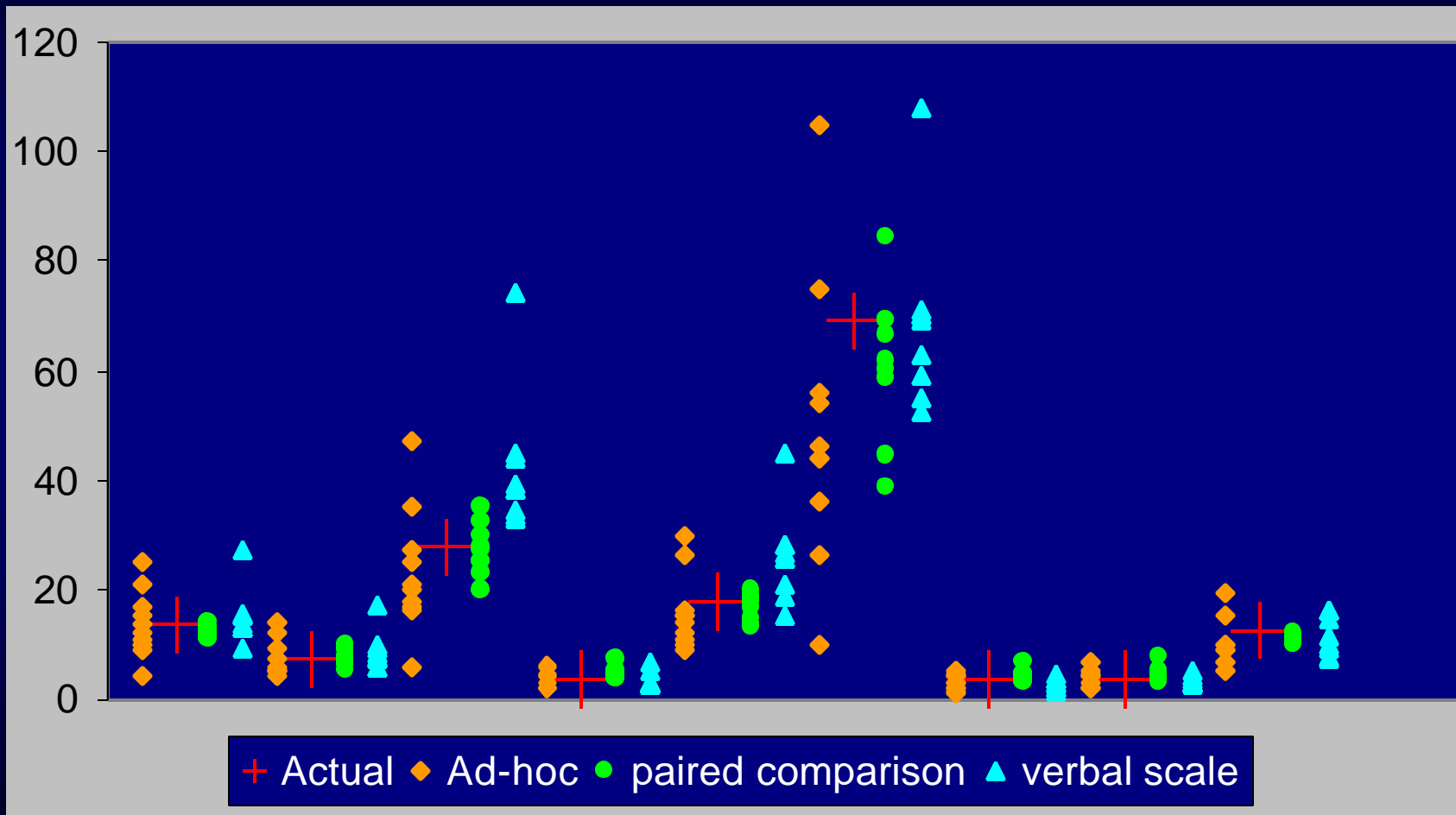
Verbal Scale

Figure	a	g	f	c	j	b
d						
a						
g						
f						
e						
j						
b						

# Evaluation - data structures



## Evaluation - geometric figures



## Evaluation - geometric figures



# Evaluation - accuracy

Are all the methods the same ?						
SUMMARY						
Groups	Count	Sum	Average	Variance		
fiw	10	370	37	157.64		
pc	9	151.7	16.855556	168.11278		
vs	7	168.3	24.042857	20.599524		
ANOVA						
Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	1981.862	2	990.93109	7.8937886	0.002454	3.42213
Within Groups	2887.259	23	125.53302			
Total	4869.122	25				

Are there any differences between PC and VS?		
	vs	pc
Mean	24.04286	16.85556
Variance	20.59952	168.1128
Observations	7	9
Hypothesized Mean Difference	0	
df	10	
t Stat	1.545675	
P(T<=t) one-tail	0.076609	
t Critical one-tail	1.812462	
P(T<=t) two-tail	0.153218	
t Critical two-tail	2.228139	

Is PC better than FIW?		
	fiw	pc
Mean	37	16.85556
Variance	157.64	168.1128
Observations	10	9
Pooled Variance	162.5684	
Hypothesized Mean Difference	0	
df	17	
t Stat	3.438601	
P(T<=t) one-tail	0.001567	
t Critical one-tail	1.739606	
P(T<=t) two-tail	0.003135	
t Critical two-tail	2.109819	

## Conclusions

- The methods have different accuracy
- The paired comparisons method shows better performance than the ad-hoc approach
- There is not significant difference between estimations using numbers and those based on the verbal scale



## Why does it work?

- Differences rather than absolute values
- Explicit paired comparisons
- Redundant values
- The result is an average of many individual observations



## Summary and further research

- The paired comparisons method provides an structured approach for doing subjective estimations which is:
  - more precise
  - more accuratethan ad-hoc approaches.
- Future research:
  - Corroboration of results
  - Validating software verbal scale
  - Application into domains others than software sizing (effort, lead time)



## References

- **G. Bozoki, An expert judgement based software sizing model, Lockheed Missiles & Space Company and Target Software**
- **J. Karlsson & K. Ryan, A Cost-Value Approach for Prioritizing Requirements, IEEE Software, September/October 1997**
- **T. Saaty, Multicriteria Decision Making: The Analytic Hierarchy Process, RWS Publications, 1996**
- **G. Crawford and C. Williams, The Analysis of subjective judgement matrices, Rand Corporation, 1985**



**The End**