

# THE FUTURE OF DATA ANALYSIS<sup>1</sup>

BY JOHN W. TUKEY

*Princeton University and Bell Telephone Laboratories*

I. General Considerations	2
1. Introduction	2
2. Special growth areas	3
3. How can new data analysis be initiated?	4
4. Sciences, mathematics, and the arts	5
5. Dangers of optimization	7
6. Why optimization?	8
7. The absence of judgment	9
8. The reflection of judgment upon theory	10
9. Teaching data analysis	11
10. Practicing data analysis	13
11. Facing uncertainty	13
II. Spotty Data	14
12. What is it?	14
13. An appropriate step forward	15
14. Trimming and Winsorizing samples	17
15. How soon should such techniques be put into service?	19
III. Spotty Data in More Complex Situations	21
16. Modified normal plotting	21
17. Automated examination	22
18. FUNOP	22
19. FUNOR-FUNOM in a two-way table	24
20. Example of use of FUNOR-FUNOM	27
IV. Multiple-Response Data	32
21. Where are we, and why?	32
22. The case of two samples	33
23. Factor analysis: the two parts	35
24. Factor analysis: regression	36
25. Factor analysis: the middle lines	36
26. Taxonomy; classification; incomplete data	38
V. Some Other Promising Areas	39
27. Stochastic-process data	39
28. Selection and screening problems	40
29. External, internal, and confounded estimates of error	41
30. The consequences of half-normal plotting	42
31. Heterogeneous data	42
32. Two samples with unequal variability	43
VI. Flexibility of Attack	44
33. Choice of modes of expression	44
34. Sizes, nomination, budgeting	45
35. A caveat about indications	46

Received July 1, 1961.

<sup>1</sup> Prepared in part in connection with research sponsored by the Army Research Office through Contract DA36-034-ORD-2297 with Princeton University. Reproduction in whole or part is permitted for any purpose of the United States Government.