Analyzing Software Measurement Data: Examples of Statistical Techniques

Normality Check

TABLE 6.1A Dataset 1

Project Effort (Months)	Project Duration (Months)	Product Size (Lines of Code)
16.7	23.0	6050
22.6	15.5	8363
32.2	14.0	13,334
3.9	9.2	5942
17.3	13.5	3315
67.7	24.5	38,988
10.1	15.2	38,614
19.3	14.7	12,762
10.6	7.7	13,510
59.5	15.0	26,500

Normality Check

TABLE 6.1B Dataset 2

Module Size	Module Fan-Out	Module Fan-In	Module Control Flow Paths	Module Faults
29	4	1	4	0
29	4	1	4	2
32	2	2	2	1
33	3	27	4	1
37	7	18	16	1
41	7	1	14	4
55	1	1	12	2
64	6	1	14	0
69	3	1	8	1
101	4	4	12	5
120	3	10	22	6
164	14	10	221	11
205	5	1	59	11
232	4	17	46	11
236	9	1	38	12
270	9	1	80	17
549	11	2	124	16

Normality Check

TABLE 6.2A Summary Statistics for Dataset 1

Statistic	Effort	Duration	Size
Mean	26.0	15.2	16,742
Median	18.3	14.8	13,048
Standard deviation	21.3	5.1	13,281

TABLE 6.2B Summary Statistics for Dataset 2

Statistic	Size	Fan-Out	Fan-In	Paths	Faults
Mean	133.3	5.6	5.8	40	5.9
Median	69	4	1	14	4.0
Standard deviation	135.6	3.5	7.9	57.0	5.8

Solution to Nonnormal Data

- Use nonparametric method
- Transform to normal data
  - E.g. Logarithmic transformation
- If true probability distribution can be identified, use that distribution

Solution to Nonnormal Data

- Use nonparametric method
- Transform to normal data
  - E.g. Logarithmic transformation
- If true probability distribution can be identified, use that distribution

T-Test

Language Desired Python	Language Desired C#
7	36
8	7
6	6
8	9
35	7
7	6
6	7