

## Verification and Cleaning of Data for On-line Opinion Survey on "January 23 Rally"

### Step 1: Late submissions

Submissions received after the 23:59 of January 28, 2005 were not processed. A total of 66 submissions were kept as preliminary valid cases.

### Step 2: Unknown personal particulars

4 cases without any personal particular were dropped, and the remaining 62 cases entered the next stage.

### Step 3: Doubtful leaflet codes

1 case with repeated leaflet codes was identified, only the last submission of each case was kept. 2 cases with invalid leaflet codes were dropped; another 3 cases with null leaflet codes were also dropped, leaving behind 56 valid cases.

### Step 4: Repetitive IP addresses

9 skeptical cases with repetitive IP addresses were identified, they were examined in terms of their time-chop, nature of their outgoing website, and content of their submissions. Finally, 0 case was confirmed to be repetitive and was eliminated from the dataset. The remaining 56 submissions comprised the final sample.

Figure 1. Verification procedures for online submissions

