

Verification and Cleaning of Data for On-line Opinion Survey on "July 1 Rally 2004"

Step 1: Late submissions

Submissions received after the 23:59 of July 4, 2004 were not processed. A total of 1,851 submissions were kept as preliminary valid cases.

Step 2: Unknown personal particulars

77 cases without any personal particular were dropped, and the remaining 1,774 cases entered the next stage.

Step 3: Doubtful leaflet codes

15 cases with repeated leaflet codes were identified, only the last submission of each case was kept. 22 cases with invalid leaflet codes were dropped; another 998 cases with null leaflet codes were also dropped, leaving behind 739 valid cases.

Step 4: Repetitive IP addresses

100 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, 3 cases were confirmed to be repetitive and were eliminated from the dataset. The remaining 736 submissions comprised the final sample.

Figure 1. Verification procedures for online submissions

