

Verification and Cleaning of Data for On-line Opinion Survey on "New Year Rally 2005"

Step 1: Late submissions

Submissions received after the 23:59 of January 6, 2005 were not processed. A total of 398 submissions were kept as preliminary valid cases.

Step 2: Unknown personal particulars

3 cases without any personal particular were dropped, and the remaining 395 cases entered the next stage.

Step 3: Doubtful leaflet codes

3 cases with repeated leaflet codes were identified, only the last submission of each case was kept. 20 cases with invalid leaflet codes were dropped; another 46 cases with null leaflet codes were also dropped, leaving behind 326 valid cases.

Step 4: Repetitive IP addresses

57 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, 1 case was confirmed to be repetitive and was eliminated from the dataset. The remaining 325 submissions comprised the final sample.

Figure 1. Verification procedures for online submissions

