Wisdom of the Crowds: Survey Results

Overview of Experiment

Five questions were asked during the introduction to the wisdom of the crowd experiment. They were as follows:

- What is the total area of India in km^2?

- In what year was the painting made?

- How many dots are in the figure?

- How many runs will India score in the upcoming cricket match?

- How many runs will Pakistan score in the upcoming cricket match?

The painting question was a multiple choice question where participants were asked to choose between four possible years. All others were open-ended where participants were asked to enter text.

Data Preprocessing and Outlier Detection

Visual inspection of the data exhibited obviously erroneous data points for two of the tasks which had to be removed prior to proceeding with the data analysis. These were mostly due to incorrect units used (eg. Lakhs instead of Km), or participants submitting answers before the images were shown and explained (eg. answers of 0 dots). For the number of dots question, observations less than 10 were removed (4 observations). For the landmass of India question, observations less than 100 were removed (1 observation).

After discarding these values, we use the standard interquartile range method to identify outliers. The interquartile range (IQR) is given by the difference between the third and first quantiles, Q_3 - Q_1. Any value that falls outside the range $\{Q_1 - 1.5 * \text{IQR}, Q_3 + 1.5 * \text{IQR}\}$ is a suspected outlier.

Summary Statistics

The summary table below represent survey results after the removal of outliers. For the mean, truncated mean and geometric mean, the percent of responses that are worse than the crowd estimate is included in parenthesis.

The histograms in Figures 1-4 show the distribution of the responses. The red line indicates the correct answer and the dotted blue line indicates the average guess of participants.

Task	Correct value	Median	Mean	25% Trunc mean	Geometric mean	Nr of responses	Min	Max	St.Dev	Skewness
Nr Dots	1, 149	587.5	812.2 (80%)	635.7 (68%)	422.5 (57%)	76	10	3, 000	769.6	1.2
Yr Painting	1890	1820	1847.8 (71%)	1848.5 (71%)	1847.8 (71%)	83	1750	1960	63.7	0.2
Runs India	300	280	282.8 (71%)	283.6 (71%)	281.9 (71%)	78	200	368	32.7	0.1
Runs Pakistan	224	256	253.2 (55%)	253.8 (55%)	251.9 (55%)	77	175	340	33.9	−0.1
Landmass India	3, 287, 590	1, 000, 000	2, 077, 340 (77%)	1, 355, 714.2 (63%)	475, 902 (41%)	71	100	10, 000, 000	2, 472, 027	1.6

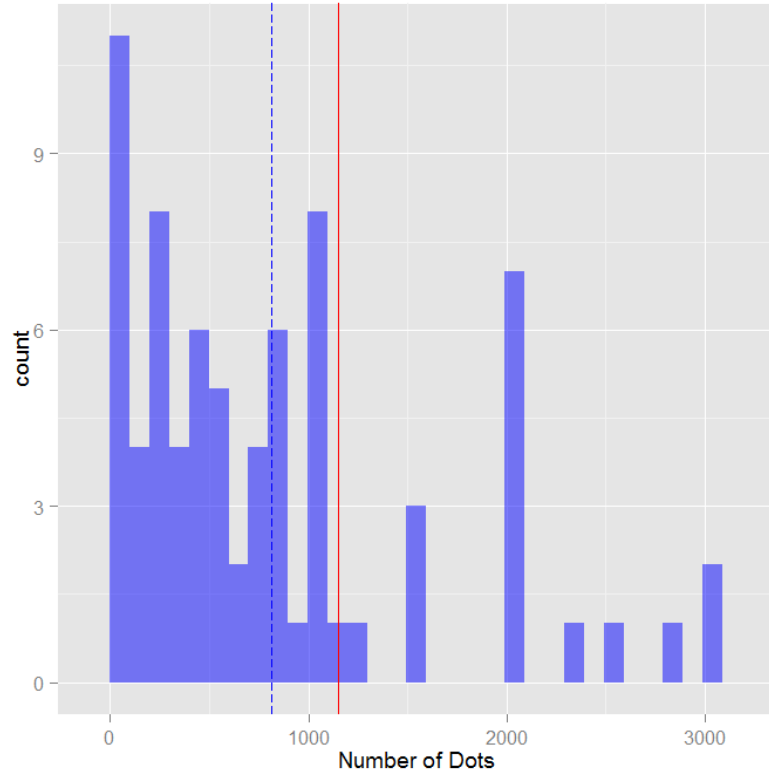Table 1: Summary Statistics

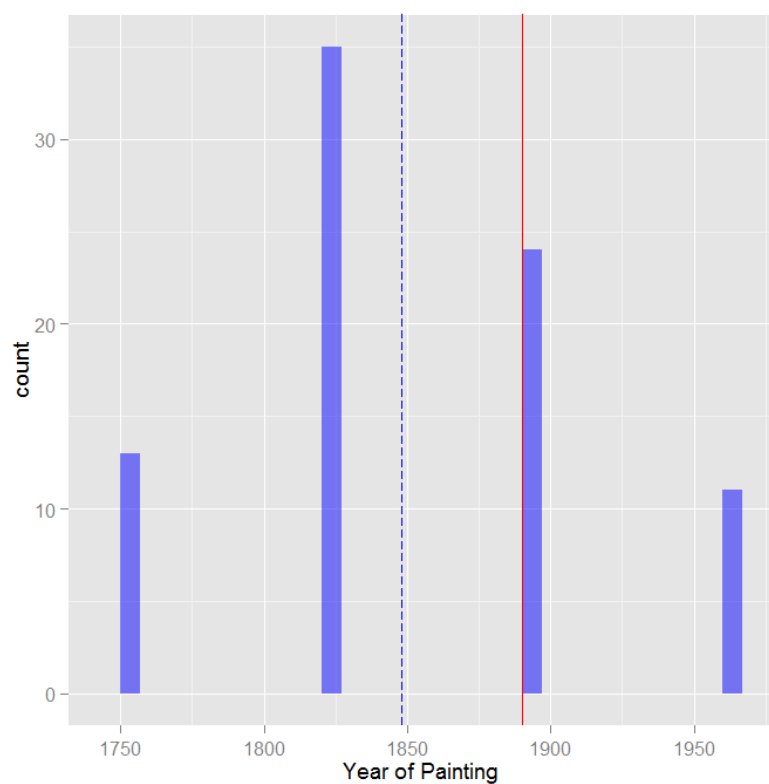Figure 1: Distribution of guesses for the number of dots

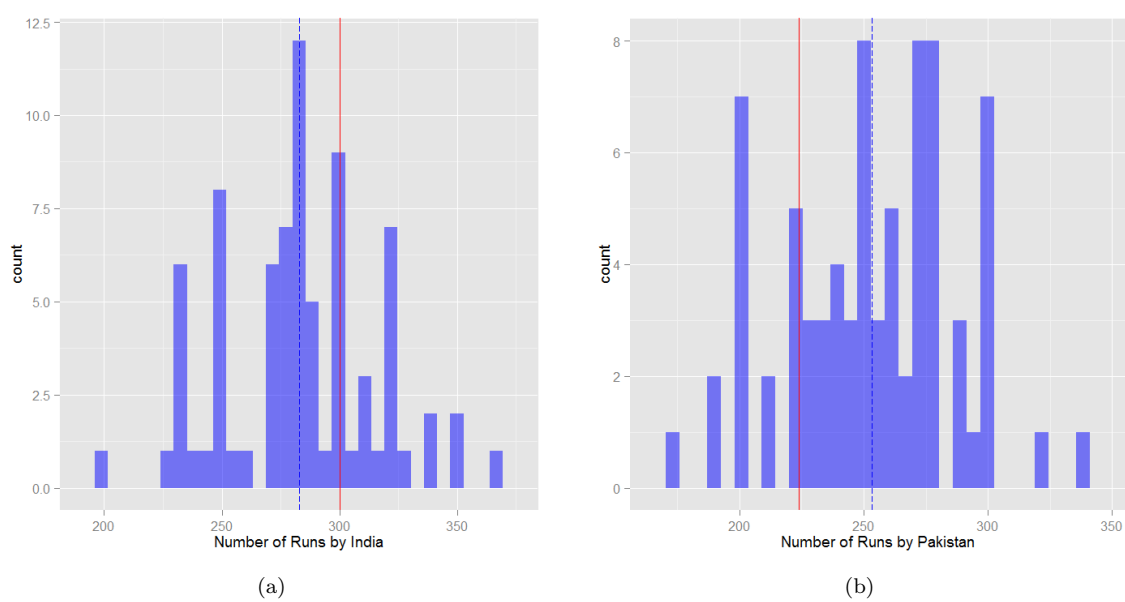Figure 2: Distribution of guesses for the year of painting

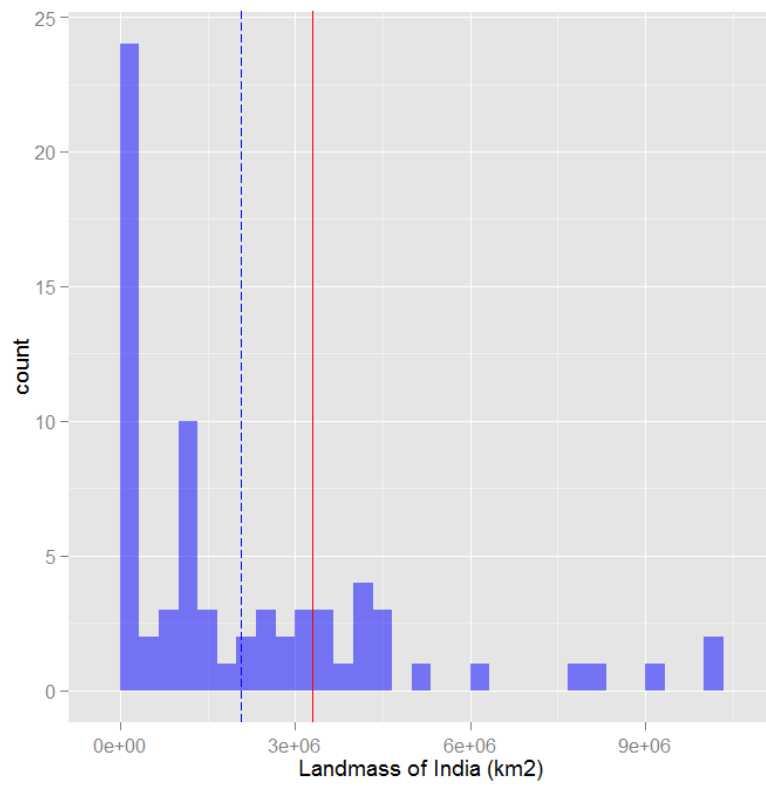Figure 3: Distribution of guesses for the number of runs for India 3(a) and Pakistan 3(b).

Figure 4: Distribution of guesses for the landmass of India