Z-scores for poll data

You are given that P_D and P_R are poll results for the Democratic and Republican candidates, with 3.5% margin of error for each candidate separately.

It follows that the poll margin $P_M = P_D - P_R$ has 7% margin of error. Let $M = D - R$ be the actual margin from the election. If the poll margin of error is correct, then

$$P(P_M - 7 \leq M \leq P_M + 7) = 0.95$$

should hold.

Now suppose we have multiple polls for the same race. Let $\bar{P}_D$ and $\bar{P}_R$ be the average poll values for the Democratic and Republican candidate, respectively, and let $\bar{P}_M = \bar{P}_D - \bar{P}_R$ be the margin between the averages. If there are k independent polls being averaged, then $\bar{P}_M$ has margin of error $7/\sqrt{k}$. Therefore

$$P(\bar{P}_M - 7/\sqrt{k} \leq M \leq \bar{P}_M + 7/\sqrt{k}) = 0.95.$$

From a simple average $\bar{X}$, we can construct a CI for EX as follows:

$$P(\bar{X} - 2 \cdot \mathrm{SD}(\bar{X}) \leq EX \leq \bar{X} + 2 \cdot \mathrm{SD}(\bar{X})) \approx 0.95.$$

By comparing the expressions for the CI, we get

$$\mathrm{SD}(P_M) = 3.5$$

$$\mathrm{SD}(\bar{P}_M) = 3.5/\sqrt{k}$$

The expected values of P_M and $\bar{P}_M$ are both assumed to be M. The Z-score measures the discrepancy between a value and its hypothesized mean, in standard deviation units. In general, if X is throught to have expected value E and variance S, then the Z-score is

$$Z = \frac{X - E}{S}.$$

If E really is the mean of X, and S is the true standard deviation of X, then Z should follow a standard normal distribution. Given one value, it is not always easy to tell whether this is the case (although if we observe, say, $Z = 5$, it is unlikely to be true). If we observe a number of values, for instance the poll results or poll averages for many races, then it becomes possible to judge whether the Z-scores actually behave like standard normal random variables. If they do not, this suggests that either the polls were biased, or the stated margins of error were wrong.