Australian Bureau of Statistics
Locked Bag 10 | Belconnen ACT 2616

Statistical analysis of results of the audit of scanned ballot papers for the 2024 ACT Legislative Assembly election

Elections ACT conducted a quality assurance audit to verify the accuracy of the count of paper ballots in the 2024 ACT Legislative Assembly election. Paper ballots are scanned to electronically capture the vote submitted on the ballot paper. The audit selects a representative sample of ballot papers and manually compares the paper ballot to the electronic capture to verify that the vote has been accurately recorded. The number of discrepancies detected in the audit sample is recorded and used to produce a statistical estimate of the accuracy of the ballot scanning process.

2024 audit sample and results

The 2024 audit was designed to use a sampling skip of 95, meaning an audit sample size of 1 in every 95 paper ballots. There is a total of 126,245 paper ballot in the population, with a final audit sample of 1,367. The sample is spread evenly across the five electorates. No errors or discrepancies were detected in any of the audited ballots.

Statistical confidence intervals for the number of errors in the full population of paper ballots

Although no scanning errors were detected in the audit sample, it is still possible that there are errors within the full population of ballot papers, but that none of the erroneous papers were selected for audit. Statistical methods are used to produce confidence intervals for the number of erroneous papers in the full population of paper ballots.

The audit sample has been selected randomly, without replacement, so that every ballot paper has the same chance of selection in the final audit sample. This can be described as a Hypergeometric distribution.

The Hypergeometric distribution describes a sampling process where a sample of size n is selected without replacement from a population of size N which contains K objects which a particular feature. For example, if a sample of n marbles is selected from a jar of N marbles that includes K red marbles, then the number of red marbles selected in the sample is a random variable, denoted by X. The Hypergeometric distribution gives the probability that X=0, or that X=1, or that X=2, and so on; or the probability that no red marbles are selected in the sample, or that 1 red marble is selected, or that 2 red marbles are selected, and so on.

Selecting a sample without replacement means that after the first selection of a marble is made, that marble is set aside and then the second marble is selected from the N-1 remaining in the jar. Then the third is selected from the N-2 remaining and so on. With replacement sampling is an alternate method where after each selection is made the marble is returned to the jar before the next selection is made. The distribution of X under with replacement sampling is the Binomial distribution. The Binomial distribution is often used as an approximation of the Hypergeometric when the size of the population, N, is very large compared to the size of the sample drawn, n. The approximation can be less accurate though when K is close to zero or close to N. As the number of erroneous ballot papers (K) may well be close to zero in this situation the Hypergeometric distribution is considered to be the best choice, although the Binomial approximation does give very similar results.

The number of erroneous ballot papers selected in the audit sample is distributed as a Hypergeometric distribution with parameters N, K, n, where

- N denotes the size of the population, with N = 126,245 (the total number of paper ballots);
- K denotes the number of erroneous ballot papers in the population (the value of K is unknown); and
- n denotes the size of the audit sample, with n = 1,367.

The probability of selecting 0 erroneous ballots under a Hypergeometric distribution has been calculated in Microsoft Excel using the HYPGEOM.DIST function. To calculate the probability of selecting 0 erroneous ballot papers in the audit sample when there are K erroneous ballot papers in the population, the excel function is HYPGEOM.DIST (0,1367, K,126245,0). The smallest integer value of K that returns a probability less than 5% is 275. This means that if there are 275 or more erroneous ballot papers in the population, then there is less than a 5% chance of selecting an audit sample with zero erroneous ballot papers.

This result is used to state that there is 95% confidence that there are 274 or fewer erroneous ballot papers in the population, or equally that the rate of erroneous ballot paper in the population is smaller than 0.22%.

Following the same method, there is 99% confidence that there are 422 or fewer erroneous ballot papers in the population, or equally that the rate of erroneous ballot paper in the population is smaller than 0.34%.