The Statistical Basis for Risk Limiting Audits

Steve Duenkel
Mason County Auditor

In an election, risk is defined as a paper ballot that is incorrectly scanned and tabulated. One error is too many. A 100% review of very scanned ballot image with every paper ballot, is not allowed within the RCW's and would be costly. We turn to statistical sampling as an alternative. Through this sampling, we want to audit the digital images of ballots to ensure that they match the paper ballot scanned (and are correctly tabulated).

For a given election, the number of ballots cast has a population of size N. The values of the population are numbers $x_1, x_2, \ldots, x_N$. For analytical purposes, a correct match between the scanned ballot image and the paper ballot = 1 (success). Any discrepancy between the scanned ballot image and the paper ballot = 0 (failure).

When we sample the total population of ballots (N), the samples are random (selected by computer) and without replacement a for specific contest. This means that the sample values are not independent. With simple random sampling without replacement, the sample values are not independent. The covariance between any two different sample values is not zero. In fact, one can show that:

Covariance between two different sample values: $\operatorname{cov}(X_i, X_j) = -\frac{\sigma^2}{N-1}$ for $i \neq j$

Where σ is the population standard deviation: $\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}x_i^2 - \mu^2}$.

This fact is used to derive these formulas for the standard deviation of the estimator and the estimated standard deviation of the estimator. The parameter of interest is usually called the population proportion, or the population mean.

Population proportion is $p = \frac{1}{N}\sum_{i=1}^{N}x_i$.

The unbiased estimator of the population (sample) proportion is $\hat{p} = \frac{1}{n}\sum_{i=1}^{n}x_i$.

Reference: *Mathematical Statistics and Data Analysis*, John A. Rice. Wadsworth, 1988, 1995.

The population mean is $\mu = \frac{1}{N}\sum_{i=1}^{N} x_i$.

The unbiased estimator of the population mean (sample mean) is $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$.

The first two columns are the parameter and the statistic which is the unbiased estimator of that parameter.

		standard deviation of the estimator	estimator of the standard deviation of the estimator
μ	$\overline{X}$	$\sqrt{\frac{\sigma^2}{n}\left(1-\frac{n-1}{N-1}\right)}$	$\sqrt{\frac{s^2}{n}\left(1-\frac{n}{N}\right)}$ where $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$
p	$\hat{p}$	$\sqrt{\left(\frac{p(1-p)}{n}\right)\left(1-\frac{n-1}{N-1}\right)}$	$\sqrt{\frac{\hat{p}\left(1-\hat{p}\right)}{n}\left(\frac{n}{n-1}\right)\left(1-\frac{n}{N}\right)}$

Population standard deviation: $\sigma = \sqrt{p(1-p)}$

We have conducted Risk Limiting Audits in specific contests for the past 7 elections to confirm that the paper ballot, the scanned image, and cast vote record are consistent. Here are the results:

Election	Ballots Cast (N)	Random Sample (n)	Ballot Discrepancies
November 4, 2021 General	17,517	70	0
February 8, 2022 Special	2.770	113	0
April 26, 2022 Special	1,413	17	0
August 2, 2022 Primary	20,140	114	0
November 8, General	29,248	62	0
February 14, 2023 Special	8,297	174	0
April 25, 2023 Special	1,334	23	0

Copies of the reports for each of these audits can be found on the Mason County Elections Website under “Previous Elections – Results & Data”.

Reference: *Mathematical Statistics and Data Analysis*, John A. Rice. Wadsworth, 1988, 1995.