Statistical Audits–Why and how much?

2008 California Association of Clerks and Election Officials
100th Anniversary Meeting
Long Beach, CA
8–11 July 2008

Philip B. Stark
Department of Statistics
University of California, Berkeley
statistics.berkeley.edu/∼stark

Vote Counting

- Counts are subject to various kinds of error.

- Counting errors $\Rightarrow$ risk of naming the wrong winner.

Audit to deter & detect fraud
monitor/improve equipment, procedures, & software
ensure total error too small to change the outcome

Explanation of the error not important for verifying outcome

Statistical Audits

Can *limit* and *quantify* that risk.

Could guarantee that, if the election is certified,

> either
>
> machines named the right winner
>
> or
>
> something really unlikely happened (say, 1 in 100)

even if the voting system has hardware or software bugs.

Selecting precincts at random is essential.

Fundamental Principle

Keep looking until you are confident that even if you looked at everything, you would not find enough error to change the winner.

Confidence depends on

1. margin (amount of error that could change the outcome)
2. how much error each precinct can hold
3. how precincts are selected for audit (*sampling design*)
4. number of precincts audited
5. error the audit finds

Complete procedure says:

- how many precincts to audit initially
- given the discrepancies in the audit sample, whether to certify or audit further
- eventually, "certify" or "full recount."

Ensures chance of certifying wrong winner is at most 1%, e.g.

Heuristics

Sample too small $\Rightarrow$ can be likely that discrepancies in the sample will be small or zero, even if machines named the wrong winner.

No look, no see: absence of evidence is not evidence of absence (of error).

Sample big enough $\Rightarrow$ likely to see big discrepancies in the sample if machines named wrong winner.

Smaller samples $\Rightarrow$ lower confidence.

Larger discrepancies in sample $\Rightarrow$ lower confidence.

Rigorous statistical audit

> If it's very likely that the audit would have found larger discrepancies than it did find, if the outcome is wrong, certify.
>
> Otherwise, keep counting.

With this approach, if an election is certified, either the correct winner was named, or something very unlikely happened.

Logistical Issues

Must commit to counts before selecting sample

Staging/stratification: cross-county, absentee, provisional, etc.

"Escalation" within strata? Across all? Exclude strata?

Other sampling schemes: lessons from financial auditing

Confidence level? Vary by race? Random sample of races? Legislative input needed.

Data, data, data: Preliminary Statement of Vote in machine-readable form.