**THE CALIFORNIA ASSOCIATION OF CLERKS AND ELECTION OFFICIALS (CACEO)**

**ELECTION COSTS PROJECT**

Welcome to the Data Analyses and Research Briefs page of the CACEO Election Costs Project! This ongoing project was made possible by a grant from the James Irvine Foundation and began in 2014. The goal was to collect actual election costs from California’s fifty-eight counties for statewide elections spanning the previous ten years. Not all counties were able to participate, and of those that did participate, not all had access to cost data going back to the earlier elections. In collecting these data, we attempted to achieve uniformity – not a small feat given differing local billing methodologies and record keeping systems. This introduction will provide information necessary to understand the series of research briefs that follow. Thank you for your interest in this project!

**Introduction to Data Analyses and Research Briefs**

The CACEO election cost data offer transparency and insights reflecting reports from county officials on election costs; the components of those costs; the electorates; and voting processes, technologies and systems used in primary, general, and special statewide elections spanning 2004 through 2014. Much of the information was entered retrospectively – as long as a decade after the respective election – and counties varied on the extent to which they identified, differentiated and tracked costs. Not all of the fifty-eight counties participated – although currently 47 counties have provided at least some information on one or more elections.

It comes as no shock that costs vary greatly - Los Angeles spent more than $30 million on its 2014 general election, nearly 4,000 times more than the $7,896 reported by Sierra County – and when comparing individual counties, potential reasons for differences between them are plentiful. Los Angeles is home to nearly 10 million residents, includes 88 distinct cities and towns in its boundaries, serves nearly 5 million registered voters, and counted about 1.5 million votes for more than 300 candidates and measures on ballots printed in 9 languages in addition to English. In contrast, the 3,127 residents of Sierra County, which includes only a single incorporated place, reflect 2,232 registered voters, 1,727 of whom voted on the 42 candidates and measures in ballots that were printed in two languages – English and Spanish.

Drawing insights from these data requires careful thought about standardizing measures. Should costs be identified per capita? Per voting age adult? Per registered voter? Per ballot cast? Per voting opportunity (the sum of candidates and measures)?  Each of these ways of standardizing can serve an analytic intent, but each highlights a different facet of costs. Once standardized, a second issue emerges. What categories of factors do we expect to affect costs or cost composition? Size is clearly a factor, but so is jurisdictional complexity – the number of cities and districts that lie within a county and, relatedly, the number of candidates and measures that are offered to voters in appropriate jurisdictions.

Complexity in the electorate also plays a role: how does the number of languages each county provides services in, the levels of linguistic isolation of members of the public, or the extent of relocation of voters – all of which may affect the accuracy of voter registration records – shape the process and costs of an election?  Counties moreover face very different costs of living and typical wage rates for their labor forces; these affect the amount spent directly on staff, other personnel, and locally produced services and supplies, and may also influence choices on relying on more or less labor-intensive processes and technologies.   Election processes and technologies are also expected to influence the level and composition of costs depending on choices of paper ballots, DREs, the relative reliance on polling places versus vote by mail ballots, and levels of challenges to ballots.

These categories of factors are examined in a set of research briefs that look at how cost and cost composition vary among counties for which these factors differ. The factors are not independent of one another, however, and the extent of correlated variation in the factors combined with the relatively small number of county participants limit strong conclusions. The briefs approach this by examining how costs differ by each set of factors for the most recent elections, how those factors co-vary, and, for a small set of counties with consistent data over time, how those costs and factors have trended.

A third issue is what costs or other outcome measures these data can illuminate. There are questions here that are interesting and useful, for example: how much was spent? What share of it was for staff? How much did ballots cost?

In approaching this dataset, we looked at what data are consistently and correctly available?  Data items become increasing less available as the level of detail increases and as the time of the election recedes further into the past, and is subject to data entry errors (e.g. entering 123456 instead of 1234.56) or differing understandings of data being *unavailable* as opposed to *not applicable*. Where possible, we correct or exclude such items, but expect some inconsistencies to remain.

If a single point is clear to us, it is that data improve by being used. The quality of underlying items improves as errors or misunderstandings are corrected, the effects of county compositional reporting by election period increases as trend analyses are run, and the understanding of associations in the data become clearer as analysts, staff and the public respond to and explore the resources and reports. We see this work as a baseline and beginning for those improvements.