ANALYSIS & INFERENCE

1489 Baltimore Pike, Suite 305
Springfield, PA 19064

P: 610-543-0159
F: 610-543-8952

Sampling Petition Signatures

Kan Qiu, Zhiming Yu, and Gang Chen *v.* Kim Wyman, in her official capacity as Secretary of State of the State of Washington

Prepared on Behalf of Plaintiffs

by

William A. Huber, PhD

March 21, 2019

Introduction

Summary of opinions

¶1 This report assesses sampling procedures used by the Secretary of State ("SoS") to certify petitions in support of Initiative I-1000 ("I-1000"). It examines two issues: (a) whether those procedures conform with the governing regulations at WAC 434-379-010 ("Random sampling procedure") and (b) how departures from the required procedures affect the determinations made by the SoS in her certification [Wyman 2019].

¶2 My examination of documents provided by the SoS leads to three principal conclusions.

a. The SoS did not carry out an "unrestricted random sample" as required by the regulations. The procedure it uses does not permit objective determination of the number of pairs of duplicate samples in a population of signatures based on a sample, which is one of its main objectives.

b. The repeated processes of sampling, modifying the population of signatures "as part of the review process" of the sample, and only partially resampling the population introduces unquantifiable, but possibly large, uncertainties in the determinations made from the sample results.

c. The number of petition signatures that could not be matched to voter registration signatures increased 20% from the first to the final sample created by the SoS. The statistical variation ordinarily exhibited by random samples cannot explain this large difference. It is likely, therefore, that at least one (and possibly both) of these samples do not reflect the populations of signatures they are intended to represent. This calls into question all inferences made by the SoS about the numbers of valid and invalid signatures in either population.

¶3 To summarize, these conclusions imply that the "protection against misrepresentation of the population" [Snedecor and Cochran] afforded by an unrestricted random sample has not been achieved. Therefore, the SoS has not shown an objectively valid basis to certify the I-1000 petition.

¶4 One notable consequence is to call into question the assertions of the Amicus brief of I-1000 sponsor Nathaniel Jackson [Ahearne 2019]. The brief argues that concerns about the counting of 218 petition sheets (bearing 4158 signatures, a bit more than one percent of the total) are "legally irrelevant" because the population size ("total of … 395,938 signatures") is larger than the number of valid signatures required for certification [*id.* at p. 3]. The foundation of this argument is the implicit assumption that most, if not all, of the signatures submitted to the SoS are valid. This assumption is unverified. My conclusions (b) and (c) raise doubts concerning the reliability of inferences made about the numbers of valid signatures based on any of the samples used by the SoS. The sample results already show that an appreciable portion of all signatures are invalid or duplicates. Flaws in the sampling procedures (conclusions (a) and (b)) and significant inconsistencies in the results (conclusion (c)) preclude an objective assessment of whether enough valid signatures are present to meet the threshold for certification. In principle, the inclusion or exclusion of even 218 petition sheets could make a difference. This appears to be a real prospect because the number of duplicates found in the first sample (25) was as close as it is possible to get to failing certification and because the projected numbers of invalid signatures have varied appreciably between samples.

Qualifications

¶5 My qualifications to carry out the work described in this report include over 30 years as a lead statistical consultant where I have been responsible for sampling and interpreting data in hundreds of varied settings, including checking the quality of data and providing expert opinions in legal disputes. I earned a PhD in mathematics from Columbia University and hold a current PStat certification ("Professional Statistician") from the American Statistical Association.

Background

¶6 On January 4, 2019, the sponsor of I-1000 submitted signed petition sheets to the SoS Elections Division. After removing 64 petition sheets with invalid text [Augino decl. at ¶16], the SoS reports having obtained a random sample of the signatures on the remaining sheets according to WAC 434-379-010 (1): "Take a minimum three percent unrestricted random sample of the signatures submitted." [Wyman 2019 number 1.] (Appendix B below reproduces the full text of WAC 434-379-010.)

¶7 Continuing to follow the nine steps prescribed by WAC 434-379-010, the SoS's examination of the sample consisted of counting invalid signatures and pairs of duplicated signatures in the sample (step 2). These counts furnished the data for calculations prescribed by steps (6) and (7) to "determine the acceptable number of pairs of signatures in the sample," followed by comparing this acceptable number to the observed number of pairs in the sample.

¶8 Relying on these comparisons [Wyman 2019 number 8], the SoS declared the "Initiative to the Legislature No. 1000 to contain sufficient signatures" [*id.*, p. 2].

Objectives

¶9 Counsel for Plaintiffs obtained supporting documentation from the SoS in the form of spreadsheets and (partially redacted) emails among SoS staff. These materials contain details of the random sampling and decision procedures conducted by the SoS. Counsel asked me to examine these materials to determine whether the sampling and decision procedures conformed with WAC 434-379-010 and, to the extent they might have departed from the requirements, what effect that might have on the SoS's declaration of sufficient [numbers of] signatures. Specifically, counsel's questions were

a. What is an unrestricted random sample as required in step one of the sampling process? Does that have specific meaning in statistics?

b. Did the office [of the SoS] use an unrestricted random sample?

c. If not, is it possible to make informed estimates of the possible range of alteration in final outcome that could result from a proper sample?

Assessment of the Sampling Procedure

Timeline

¶10 Augino (2019a) ("Timeline") establishes that the SoS sampled the petitions at least twice[1]:

a. A sample of 11,881 signatures from a population of 394,716 signatures was created and evaluated between January 16 and January 25, 2019. The SoS found "25 pairs of duplicate signatures" in this sample.

b. During review, the SoS removed 61 petition sheets from the population and added a box of 141 sheets to the population. (Sheets typically bear 20 signatures each.)

c. A second sample of 11,919 signatures from the increased population of 395,981 signatures was created and evaluated between January 31 and February 4.

d. During review, the SoS discovered "three additional incorrect sheets" (with 43 signatures) in the population. The SoS checked that the random sample did not include any signatures from these sheets.

[1] Table 1 at ¶29 summarizes the timeline.

e. Only 13 pairs of duplicate samples were found in this final sample of a population now reduced to 395,938 signatures.

Supporting documents

¶11 Three Excel files, provided by or originating with the work of the SoS, contain details of the foregoing sampling efforts. I describe the contents of these files in Appendix A below. These files show the following:

a. Signatures were identified by listing the petition sheets by page number and box number in an apparently arbitrary[2] order: the "sampling frame." Implicitly, they were placed in a list and numbered 1, 2, 3, and so on.

b. In all sampling efforts, a common set of random numbers was used to select signatures. These numbers were created by randomly generating whole values between 1 and 66, inclusive, which represent the number of lines to skip from one sampled signature to the next in the sampling frame: I call these the sampling intervals. Values between 1 and 65, inclusive, appear to have had equal probabilities of being selected, while 66 had less chance of being selected than the others[3].

c. The samples differ due to differences in the sets of pages included in the sampling frame. Such differences caused the identifiers of some (but not all) of the signatures to change each time the population was changed, because as some pages were deleted and others were inserted, the positions of signatures further down in the sampling frame changed accordingly.

---

[2] "Arbitrary" does not mean "random." The order was established by how the signatures were grouped on their pages, how those pages were put into boxes, and the order in which the boxes were identified, all of which are non-random characteristics and could be related to patterns of occurrence of invalid signatures and duplicate signatures. However, in a truly unrestricted random sample, the order in which members of the population are listed in the sampling frame is immaterial. In all other kinds of sampling, the order can matter.

[3] This difference in chances is not inherently problematic. It was required by (a) the need, stipulated at WAC 434-379-010 (1), to select 3% of the population; and (b) an apparent desire to *restrict* intervals to be as small as possible. Signatures must therefore be selected *on average* once for every 100/3 = 33.33 lines of the sampling frame. That average was achieved by letting the intervals vary from 1 through 2(33.33) – 1 = 65.67. Because that is impossible to do—65.67 is not a whole number—it was approximated by sometimes generating random intervals of 66, but at a frequency only 0.67 times as great as any shorter interval.

d. The acceptable number of duplicate signatures in any sample (that is, the largest number that can justify certifying the petition) depends on the sample size and the population size. However, because variations in those sizes were slight from one sampling effort to the next, the acceptable number was always 25 duplicates[4].

Interpretation of the regulations

¶12 The regulations [WAC 434-379-010] do not define "unrestricted random sample." I understand this phrase to refer to what is more commonly known as a "(simple) random sample without replacement," as described in a widely adopted handbook of statistical methods:

> *In one form of random sampling, each member of the population has an equal chance of being selected at every draw. … We can select a random sample by having discs numbered from 1 to N representing the members of the population. The discs are put in a receptacle, mixed thoroughly and one sample member is drawn out. After its number as been noted we replace it, mix again, draw a second member, and so on. This method is called random sampling* with replacement. …
>
> *A variant of this method of sampling is to allow any member of the population to appear only once in the sample. … [S]ince the purpose of the sample is to provide information about the population, it makes sense to have as many* distinct *members of the population as possible in the sample. …*
>
> *The intuitive appeal of random sampling lies in its fairness—it gives every item in the population an equal chance of being selected and measured and thus should protect us against distortion or misrepresentation of the population.*
>
> [Snedecor and Cochran section 1.6, pages 6 – 7. Emphases in the original.]

---

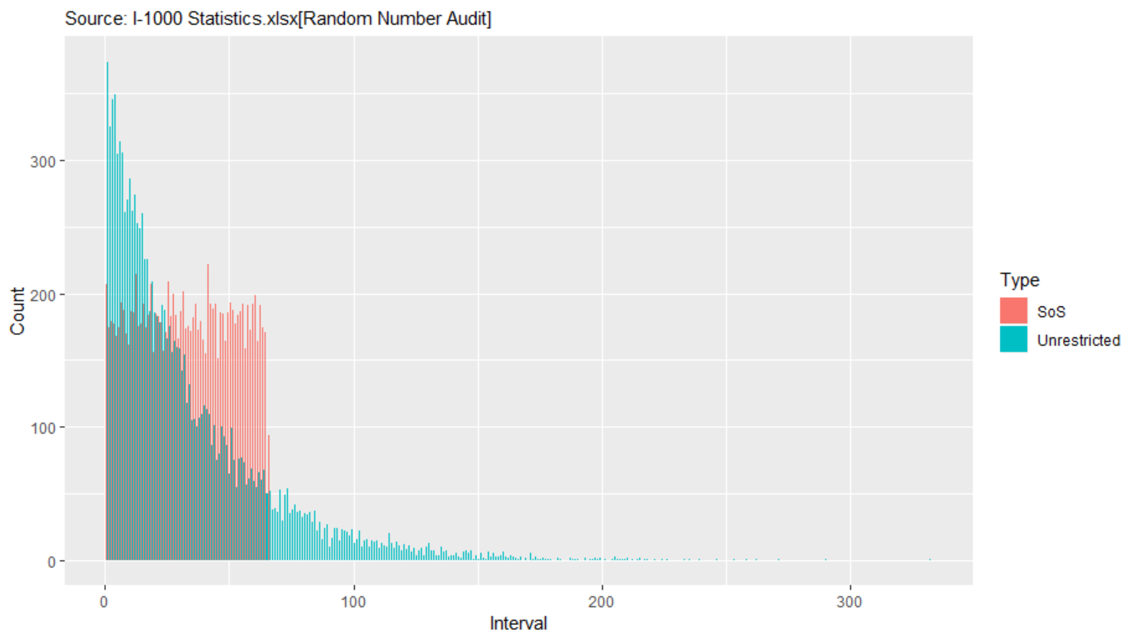
[4] This is a vague upper limit: when the count of duplicate signatures in the sample is "greater than" the acceptable number, then "each signature [in the population of all signatures] shall be canvassed to determine the exact number of valid signatures" [WAC 434-379-010 (8).] When the duplicate count is "less than" the acceptable number, then "the petition shall be deemed to contain sufficient signatures" [id. (9)]. The regulations do not stipulate what action to take when the number of duplicates in the sample equals the acceptable number. However, the "acceptable number" is the result of a calculation and rarely equals a whole value. In this case it is 25.3. Thus, a duplicate count of 25 can be taken to be *less than* the acceptable number.

Assessment of the SoS Sampling Procedure

¶13 **In the sense of fairness as described by Snedecor and Cochran, the SoS samples are not "unrestricted."** This can be seen clearly in the intervals between the sampled signatures. In unrestricted random sampling, smaller intervals will occur much more frequently than larger ones, but very large intervals also occur, albeit infrequently.

¶14 **Figure 1 contrasts the distribution of SoS sample intervals to that of intervals expected to appear in any unrestricted random sample of the same population.** The red bars for "SoS" display the frequencies of each interval used in the SoS samples. The number of times each interval (of size 1, 2, 3, …, through 66) was used in the sample is depicted by the height of the corresponding bar. This bar chart reproduces the histogram created by the SoS in "I-1000 Statistics.xlsx" (see Appendix A below). The light blue bars display the frequencies of a typical unrestricted random sample of the same size. Because intervals much larger than 66 will appear in an unrestricted sample, the plot of blue bars extends further to the right than the plot of red bars. Although the blue bars depict intervals in one particular random sample, intervals in any other unrestricted random sample of this population will exhibit very nearly the same histogram.

¶15 **Figure 1: Intervals Between Signatures for Two Types of Random Sampling.**



¶16 **The two histograms of Figure 1 exhibit pronounced differences.** In particular,

a. the frequencies of intervals less than 20 are greater for the unrestricted sample and

b. the unrestricted sample exhibits many intervals longer than 66, which is the longest used by the SoS.

¶17 For the purpose of estimating how many duplicate signatures are in the population, an unrestricted random sample will be *fair* in the sense that every *pair* of signatures in the population will have an equal chance of being selected and identified and thus should protect us against distortion or misrepresentation of the population (to paraphrase [Snedecor and Cochran], *op. cit.*).

¶18 **The samples obtained by the SoS are not fair in this sense.** As Figure 1 shows, their sampling intervals are too concentrated on values between 20 and 65 compared to a fair, unrestricted random sample.

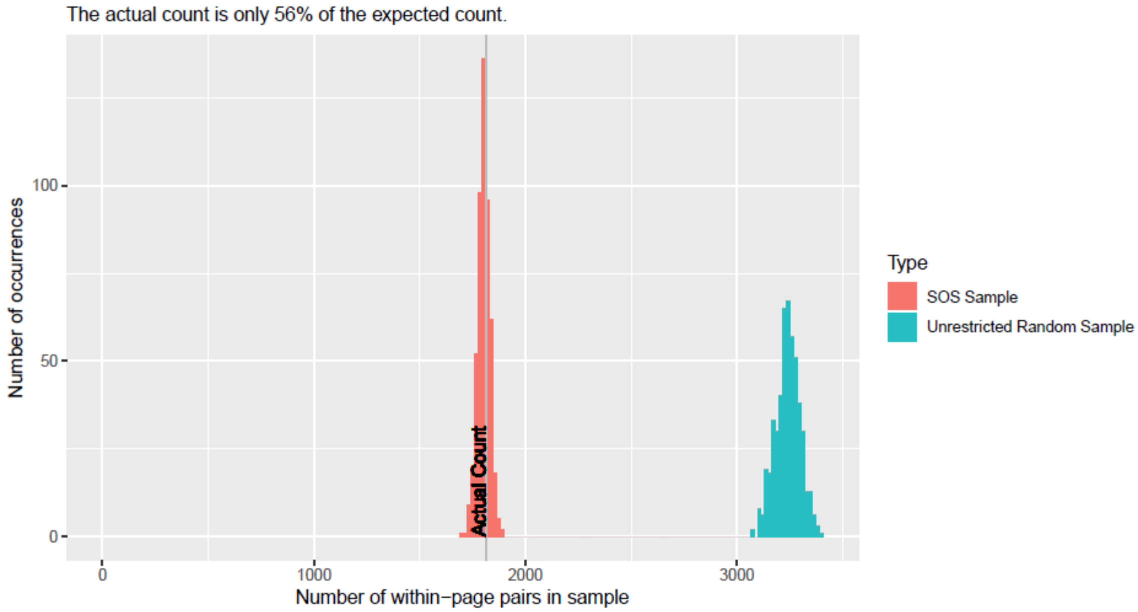
¶19 A compelling reason to demand fairness of a sample lies in the prospect that pairs of duplicate signatures in the population might result from deliberate or, at least, non-random actions. For instance, some petition sheets may exhibit many duplicates while most may exhibit none. Or duplicates might never occur within a single sheet, yet appear amongst different sheets. Thus, a fair sample should not tend to collect too many or too few pairs of signatures within common sheets. (The formula of WAC 434-379-010 (7) relies on this assumption.) The question of the *amount* of unfairness in the SoS sampling procedure arises: would it be reasonable to construe it as being "sufficiently close" to an unrestricted sample "for practical purposes"?

¶20 To study this, I used a computer to generate samples from the population described in "I-1000 Statistics.xlsx" (see Appendix A below)[5]. It generated 500 independent samples using the SoS method and another 500 independent samples using unrestricted sampling. In each of these 1,000 samples the computer found and counted all pairs of signatures that were selected from the same page[6]. Figure 2 summarizes the pair counts using a separate histogram for each type of sample.

[5] Although this population differs from the one ultimately sampled by the elimination of three pages (out of more than 20,000 pages), the difference is inconsequential.

[6] Thus, a page with just one signature in the sample contributes zero pairs, a page with exactly two signatures in the sample contributes one pair, a page with exactly three signatures in the sample contributes *three* pairs, a page with exactly four signatures in the sample contributes *six* pairs, and so on. A page with exactly k signatures in the sample contributes $k(k-1)/2$ pairs.

¶21 Figure 2: Summary of Pair Counts for Two Types of Random Sampling.



¶22 The sample first obtained by the SoS contains 1,817 within-page pairs of signatures[7]. Figure 2 displays this count with a vertical gray line. The narrow (red) histogram surrounding this value shows that the SoS sampling procedure is extremely likely to produce a comparable count: certainly one that is less than 2000. Thus this value of 1,817 is characteristic of the sampling procedure itself.

¶23 The unrestricted random samples all exhibit within-page pair counts greater than 3000, shown with the light blue histogram in Figure 2. These counts average 3240. On average, then**, the SoS procedure produces within-page pairs at less than 60% of the rate it is required to do so.** In this sense its samples are neither "unrestricted" nor are they fair.

¶24 **Because the SoS procedure is not unrestricted, the calculations of acceptable numbers of duplicate pairs required by WAC 434-379-010 (7) do not apply.** Indeed, a comparable calculation for the SoS procedure is not possible because it would depend on assumptions made about where in the population duplicates are likely to appear. (Such assumptions are unnecessary for an unrestricted sample.) Therefore, the SoS has no statistically valid basis to defend the comparison of the duplicate signature count (found in the sample) to the "acceptable" number of pairs as required by WAC 434-379-010 (8) and (9).

[7] The second and third samples also contain nearly the same numbers of within-page pairs.

Consequences of the Repeated Sampling for the I-1000 petition

¶25 With unrestricted sampling, changes to the sampling frame are easy to handle: when signatures have to be removed *for reasons not discovered by the sampling process*, any of those signatures that happen to be in the sample may simply be dropped. When new signatures are added to the sampling frame (for any reason) *and are believed to be no different than those originally in the frame*, then they can be sampled separately and the signatures in this separate sample can be adjoined to the existing sample.

¶26 The restricted nature of the SoS sampling procedure caused the SoS to implement *ad hoc* adjustments to the original sample. These adjustments affected only part of the sample: the samples have many signatures in common. Nevertheless, the count of duplicate pairs fluctuated wildly. Originally it was 25 – exactly equal to the acceptable number – and then became 13, even when only part of the sample was changed. This shows how exquisitely sensitive the count of duplicate pairs can be to the choice of sample, emphasizing the need to follow the requirement for an unrestricted sample rigorously.

¶27 **Of greater concern is the possibility that the SoS did not respond correctly to information in the original sample.** If the duplicates found in that sample led, either directly or indirectly, to the discovery of groups of sheets that had to be removed from the population ("as part of the review process" of the first sample, [Augino 2019a at p. 5]), then the validity of all inferences drawn from that sample, or any part thereof, becomes questionable. In effect, this process of sampling, discovering batches of invalid sheets, removing the sheets from the population, and then resampling just part of the population is tantamount to an iterative way of discovering problems in the population and removing them *until the* modified population *is "acceptable."* In effect, the process followed by the SoS may have acted as a mechanism to *adjust the population to the sample* rather than using the sample to characterize the population, as intended. This is not statistically valid, because it is no longer possible to draw any objective connection between statistical properties of the sample and corresponding properties of the population.

Assessment of Differences Between Two Samples

¶28 In a February 6, 2019 e-mail copied to the SoS, Lori Augino (Director of Elections) provided "the most up-to-date timeline of events with I-1000" [Augino 2019a]. This timeline documents two sampling events (which I believe are the first and last one attempted by the SoS) and summarizes how the sampled signatures were classified. The following table summarizes these data.

¶29 **Table 1: Summary of Two Sampling Events**

	First Sample	**Second sample**
Population size	394,716	395,981
Sample size	11,881	11,919
Sample outcomes		
Accepted signatures	9,150	9,047
Registrations not found	1,964	1,973
Signatures not in database	10	9
Did not match voter registration signature	732	877
Duplicate pairs	25	13

¶30 The second population differed from the first by (i) excluding 61 sheets containing 891 signatures; (ii) adding 141 more sheets containing 2,156 signatures; and (iii) removing three sheets comprising 43 signatures. These changes affected at most (891+2156+43)/394716, or 0.78%, of the original population of signatures.

¶31 The different outcomes for these two samples ought to reflect only the combined effects of (a) the operation of chance in the random selection of the first sample; (b) the operation of chance in the random selection of the second sample; and (c) the change in population.

¶32 The effects of chance in both samples can be evaluated using a standard, widely used "chi-squared" test [Snedecor and Cochran]. This test evaluates whether tables of mutually exclusive sampling outcomes, such as the "Sample Outcomes" portion of Table 1, reflect independent samples of the same population. Although the conventional chi-squared test does not account for sampling *without* replacement, the sample sizes in Table 1 are small enough compared to the population sizes that the chi-squared test will provide accurate results regardless.

¶33 Like all statistical hypothesis tests, the chi-squared test computes a "p-value." The p-value is based on a preliminary characterization of how consistent any possible table of results is with the assumption that the two samples are independent and random. The test evaluates the plausibility of this assumption—the "null hypothesis"—by supposing, hypothetically, that it is correct. This permits one to compute the chance that the data could have turned out to be *less* consistent with the null hypothesis than the observed data (in Table 1) are. This hypothetical chance is the p-value. When it is large, or at least not too small, the null hypothesis remains plausible. When the p-value is small, it constitutes evidence against the null hypothesis. Conventionally, and often in legal cases, p-values smaller than 5% or 1% are taken to be "statistically significant" evidence that the null hypothesis may be incorrect.

¶34 The chi-squared test of the sample outcomes in Table 1 has a p-value of 1/633, or 0.16%. This is due primarily to the discrepancy in counts of unmatched voter registration signatures, which increased by 20% from 732 to 877, even though the sample size hardly changed. The discrepancy between the columns of Table 1, as measured by the chi-squared statistic, will be exceeded in truly random sampling of a common population in only one out of every 633 pairs of random samples. Because this is such a small chance, it provides strong evidence that **the differences in sample outcomes are not due to the random selection of signatures**.

¶35 I investigated the extent to which this significant difference in results could be attributed to the changes in the populations of signatures. Whether each sample was obtained either as an unrestricted sample without replacement, as required, or using the restricted method employed by the SoS, it is intuitively obvious that the effect of a 0.78% change in population should create *approximately* a 0.78% change in outcomes. In particular, the expected change in the count of unmatched voter registration signatures ought to be around 0.78% of an average of 732 and 877 (the counts in the two samples), which is equal to 6.3. The variation around this expected difference due to random sampling is relatively small: with 99% probability it will not exceed 13.

¶36 I therefore re-ran the chi-squared test after bringing the counts of unmatched signatures closer together by 13. This was done by subtracting 13 from the count of 877 in the second sample and adding 13 to the count of accepted signatures in that sample. The new p-value is 1/229, or 0.44%. It is still small and continues to indicate a significant difference. Therefore, **the difference between the two samples cannot plausibly be attributed to the change in population**.

¶37 **I conclude that one of these samples, or possibly both, was not actually random and therefore does not represent the population from which it was drawn.** There are many possible ways in which the sample outcomes could have changed, such as differences in how the petition reviewers determined matches with voter registration signatures. The foregoing examination of the available data cannot not provide an explanation: it only demonstrates that *some* difference in sampling procedures occurred that was not likely due to the random mechanism used to select signatures for the samples.

¶38 I understand that the second sample was independent of the first in the sense that a different set of random numbers was generated to obtain it. If the two samples indeed shared some random numbers, that would reduce the expected amount of variation in their results. Accordingly, the differences in unmatched voter signatures seen in Table 1 would become even more unusual than my analysis indicates, thereby strengthening its conclusions.

References

Ahearne, Thomas (2019). Amicus Brief of Initiative 1000 Sponsor Nathaniel Jackson. March 19, 2019.

Augino (2019a). *I-1000 Timeline*. Email to Sheri Nelson (Deputy Secretary of State), February 6, 2019 2:57 pm.

Augino (2019b). *Declaration of Lori Augino, Director of Elections*. Qiu *et al.* v. Kim Wyman, March 4, 2019.

Snedecor, GW and WG Cochran, *Statistical Methods, Eighth Edition*. Iowa State University Press, 1989.

Wyman (2019). *Certification of Initiative to the Legislature No. 1000*. Washington State Elections Division, February 6, 2019.

Appendix A: Excel Workbooks

¶39 "I-1000 Statistics.xlsx," provided by the SoS, consists of seven worksheets. They proceed logically through identification of which signatures to sample, analysis of the sample results, and review of the work, as follows:

a. **Step 1 – Signature Count**. This enumerates every petition sheet within 108 boxes (identified as "Box 1," "Box 2" and so on and presented in that order). It lists how many signatures are on each sheet, and—by establishing a fixed sequence of boxes and sheets within each box—implicitly assigns a sequence of numbers from 1 through 395,891 to the signatures. The order in which the boxes are placed, as well as the number of pages within each box, appears to be arbitrary but not random.

b. **Step 2 – Statistics**. This sheet lists some statistical properties of (a) the population described in [Step 1 – Signature Count] and (b) the outcome of the SoS review of the sample, which determined that 2,384 of 11,919 sampled signatures were invalid. The count of 11,919 agrees with that of the second sample described in the Timeline.

c. **Step 3 – Random Sample**. In the column headed "Random Number," this sheet lists 11,919 whole numbers, all lying between 1 and 66, inclusive. From this sequence it derives a *cumulative sum* shown in a column headed "Total." A cumulative sum is the process of successively adding each new number to the sum of those preceding it. To illustrate, the random number sequence begins 61, 35, 55, 49, … and its cumulative sums are 61, 61+35=96, 96+55=151, 151+49=200, and so on.

d. **Step 3 – Random Sample (graphic)**. This sheet also includes two columns "Bin" and "Frequency," containing 66 rows of nonzero data, along with a bar chart of the frequency values titled "Histogram for I-1000". The frequency counts the occurrences of each value in the "Random Number" column on this sheet. The frequencies range from 151 through 222, except for random number 66, which occurs only 94 times.

e. **Step 4 – Validation Check**. This appears to document the SoS review of the sample. It represents that 2,872 signatures were "rejected" and that 13 were "duplicate."

f. **Step 5 – Final Report**. This sheet calculates acceptable counts of invalid and duplicate signatures as required by WAC 434-379-010. It indicates 2,859 signatures were invalid. Comparison to the numbers 2,872 and 13 shown in the preceding sheet indicates that 2,859 is the number of those signatures that were invalid outright, to which another 13 (one of each pair of duplicates) have been added to give 2,872.

g. **Random Number Audit**. The first six columns of this sheet indicate its contents were created for Initiative 1000 on 1/20/2019 at 3:58:31 pm by Christopher Donald using a "cryptographic seed." It has 11,919 data rows. The remaining columns of this sheet list "Seed used," "Random number generated," "Additional value used," "Raw result", and "Rounded result" for each row. The "Rounded result" column is identical to the "Random Number" column in [Step 3 – Random Sample].

h. **Random Number Match**. The first five columns of this sheet identify which signature corresponds to each of the cumulative values in [Step 3 – Random Sample]. The last two columns summarize the numbers of sampled signatures taken from each box of petitions.

¶40 "I-1000 Statistics_Final.xlsx," provided by the SoS, contains the first five sheets in "I-1000 Statistics.xlsx," but the numbers have slightly changed. It documents a population of 395,938 (computed by manually subtracting 43 from 395,981) and a sample size of 11,919. The enumeration and labeling of box contents in [Step 1 – Signature Count] remain the same as before. The "Random Numbers" of [Step 3 – Random Sample] are the same. The values in [Step 5 – Final Report] are consistent with removing 43 signatures from the population, without changing the sample, as described at ¶10d above.

¶41 "I-1000_Resampling.xlsx," provided by Plaintiff, consists of a single sheet formatted as two panels of data.

a. The left panel enumerates and identifies boxes and pages within boxes, as in the [Step 1 – Signature Count] sheets in the preceding two workbooks. To it is appended a "Total signatures adjusted" column. Comparing this to the original "Total signatures" column reveals that the signature counts have been set to zero for Box 61 page 110 (originally 19 signatures), Box 68 page 181 (originally 16 signatures), and Box 99 page 21 (originally 8 signatures), thereby decreasing the total number of signatures by 43.

b. The right panel reproduces the [Random Number Match] sheet of "I-1000 Statistics.xlsx." To it are appended three columns in which the box identifiers, page numbers, and signature line numbers have been "adjusted." The adjustments begin after an entry for Box 61, page 109. They reflect the removal of the foregoing three pages and 43 signatures from the population.

Appendix B: WAC 434-379-010

Random sampling procedure.

In the verification of signatures on initiative and referendum petitions, under RCW 29A.72.230, the following statistical test may be employed:

(1) Take a minimum three percent unrestricted random sample of the signatures submitted;

(2) Check each signature sampled to determine the number of valid signatures in the sample, the number of signatures in the sample which are invalid because the individual signing is not registered or the signature is improper in form, and the number of signatures which are duplicated in the sample;

(3) Calculate an allowance for the chance error of sampling by multiplying the square root of the number of invalid signatures in the sample by 1.5;

(4) Estimate the upper limit of the number of signatures in the population which are invalid by dividing the sum of the invalid signatures in the sample and the allowance for the chance error of sampling by the sampling ratio, i.e. the number of signatures sampled divided by the number of signatures submitted;

(5) Determine the maximum allowable number of pairs of signatures in the population by subtracting the sum of the number of signatures required by Article II, Section 1 of the Washington state Constitution and the estimate of the upper limit of the number of invalid signatures in the population from the number of signatures submitted;

(6) Determine the expected number of pairs of signatures in the sample by multiplying the square of the sampling ratio by the maximum allowable number of pairs of signatures in the population;

(7) Determine the acceptable number of pairs of signatures in the sample by subtracting 1.65 times the square root of the expected number of pairs of signatures in the sample from the expected number of pairs of signatures in the sample;

(8) If the number of pairs of signatures in the sample is greater than the acceptable number of pairs of signatures in the sample, each signature shall be canvassed to determine the exact number of valid signatures;

(9) If the number of pairs of signatures in the sample is less than the acceptable number of pairs of signatures in the sample, the petition shall be deemed to contain sufficient signatures and the serial number and ballot title shall be certified to the state legislature as provided in RCW 29A.72.230 or to the county auditors as provided in RCW 29A.72.250.

Source: https://apps.leg.wa.gov/wac/default.aspx?cite=434-379-010, accessed 3/14/2019.

Contents