APPENDIX 1

(165-014-0030)

Sampling to validate state petitions

The statistical formulas for determining the qualification of a petition for the ballot depend on sampling estimates for the number of distinct valid signatures in the petition.

For describing the verification methods, it is convenient to list some general notation for counts from a petition and sample.

Petition:

$N =$ number of signatures submitted (petition size)

$M =$ number of distinct valid signatures (number of electors that submit valid signatures)

$Y =$ number of valid signatures

$D =$ number of duplicates of valid signatures

$R =$ required number of distinct valid signatures

Sample:

$n =$ number of signatures selected in sample (sample size)

$y =$ number of valid signatures in sample

$e_2 =$ number of electors with two valid signatures in sample

$e_3 =$ number of electors with three valid signatures in sample

In general, e_k, will represent the number of electors with k $(k = 1, 2, 3, 4, \cdots)$ valid signatures in the sample.

The total number of distinct valid signatures in the petition is given by

$$M = Y - D, \qquad \text{(Equation 1)}$$

where the subtraction eliminates the duplicates of valid signatures. A statistical estimate for the number of distinct valid signatures can be obtained by substituting estimates for the numbers of valid signatures and duplicates of valid signatures in Equation 1. The task is to determine, from statistical estimates, whether the total number of distinct valid signatures in the petition attains the required number, $M \geq R$.

The signature verifications are made in stages. First, a random sample of 1,000 signatures is verified. If the petition does not qualify from the first sample, then the second larger random sample is verified. Qualification of the petition is then based on an estimate of M made from the combined first and second samples. In the event that the petition is not qualified from the combined sample and a second submission is made as permitted by ORS 250.105(3),a random sample of signatures from the second submission is verified. Qualification of the petition after verification of the sample from the second submission is based on an estimate of M made from the samples in the first and second submissions. The methods for determining whether a petition has a sufficient number of valid signatures are described for the following three cases: after verification of the first sample, after verification of the second sample, and, when applicable, after verification of the sample from the second submission. For each of these three cases, examples are given to illustrate the numerical calculations and conclusions.

Verification of the first sample

A petition is qualified for the ballot after verification of the first sample only if there is a high level of confidence that the petition contains the required number of distinct valid signatures. This will correspond to comparing a lower limit for the unknown number of distinct valid signatures in the petition, $M \geq M_{\mathrm{L}}$, to the required number R. The lower limit M_{L} can be obtained by substituting into Equation 1 a statistical 95% level lower confidence limit for the number of valid signatures, $Y \geq Y_{\mathrm{L}}$, and a specified upper limit for the number of duplicates of valid signatures, $D \leq D_{\mathrm{U}}$, that corresponds to an assumed upper limit for the duplication rate of at least 8%, $D_u \geq 0.08N$. An upper limit for the number of duplicates is assumed for qualification from the first sample as required by ORS 250.105(5)

Steps for determining the result from the first sample

1. Calculate the estimate for the number of valid signatures in the petition by multiplying the number of valid signatures in the first sample, $y = y_1$, by the expansion factor calculated as the ratio of the petition size, N, over the first sample size, $n = n_1$,

$$\hat{Y}_1 = \left(\frac{N}{n_1} \right) y_1 \qquad \text{(Equation 2)}$$

2. Calculate the margin of error for the estimate $\hat{Y}_1$

$$\text{MOE} = 1.645 \sqrt{\left\{ \frac{N(N-n_1)}{n_1 - 1} \right\} \left(\frac{y_1}{n_1} \right) \left(1 - \frac{y_1}{n_1} \right)} \qquad \text{(Equation 3)}$$

where the constant 1.645 is the normal deviate corresponding to the confidence probability 0.95. Next, calculate the 95 % lower confidence limit for the number of valid signatures in the petition

$$Y_{\text{L}} = \hat{Y}_1 - \text{MOE} \qquad \text{(Equation 4)}$$

3. Calculate the upper limit for the number of duplicates of valid signatures by multiplying the assumed upper limit for the duplication rate by the petition size. For example, corresponding to the 8% upper limit,

$$D_{\text{U}} = 0.08N \qquad \text{(Equation 5)}$$

4. Substitute the lower limit in step 2 and upper limit in step 3 in Equation 1 to produce the lower confidence limit for the number of distinct valid signatures in the petition

$$M_{\text{L}} = Y_{\text{L}} - D_{\text{U}} \qquad \text{(Equation 6)}$$

5. Compare the lower confidence limit, M_{L}, to the required number, R. If M_{L} is greater than or equal to R, the petition qualifies for the ballot. If M_{L} is less than R, the second sample must be verified.

Verification of the second sample

For petitions that require verification of a second sample, data are combined from the first and second samples. Calculation of the estimates for the numbers of valid signatures, duplicates, and distinct valid signatures are described respectively in the steps 1, 2, 3 for determining the qualification result from the combined sample.

Steps for determining the result from the combined sample

1. First, calculate the total number of valid signatures in the combined sample by adding the number of valid signatures in the second sample, y_2, to that of the first sample

$$y = y_1 + y_2 \qquad \text{(Equation 7)}$$

Similarly, calculate the total combined sample size

$$n = n_1 + n_2 \qquad \text{(Equation 8)}$$

Then calculate the estimate for the number of valid signatures in the petition by multiplying the number of valid signatures in the combined sample by the expansion factor calculated as the ratio of the petition size over the combined sample size

$$\hat{Y} = \left(\frac{N}{n}\right) y \qquad \text{(Equation 9)}$$

2. Determine the numbers of electors in the combined sample with two valid signatures, e_2, and three valid signatures, e_3. Then calculate the estimate for the number of duplicates

$$\hat{D} = \left(\frac{N}{n}\right)^2 \left(e_2 + e_3\right) + \left(\frac{N}{n}\right) e_3, \qquad \text{(Equation 10)}$$

where the expansion factor for the total number of electors with either two or three valid signatures in the sample, $e_2 + e_3$, is calculated as the square of expansion factor for y in Equation 9 and the expansion factor for e_3 alone is the same as that for y in Equation 9.

The purpose of the last component of the sum in Equation 10 is to cancel the contribution of the third valid signature in Equation 9 to the estimate for the number of distinct valid signatures in Equation 11. Then the net contribution of zero for the third signature is the same as for an invalid signature. The contribution of the second valid signature to the estimate of duplicates is accounted for in the first component of the sum in Equation 10. If the extremely rare event were to occur where four, or more, valid signatures from an elector are found in the combined sample then Equation 10 can be generalized so that the forth, or additional, signatures contribute zero to the estimate for the number of distinct valid signatures.

3. Substitute the estimates from steps 2 and 3 in Equation 1 to give the estimate for the number of distinct valid signatures in the petition:

$$\hat{M} = \hat{Y} - \hat{D} \qquad \text{(Equation 11)}$$

4. Compare the estimate $\hat{M}$ to the required number, R. If $\hat{M}$ is greater than or equal to R, the petition qualifies for the ballot. If $\hat{M}$ is less than R, the petition does not qualify for the ballot.

Verification of the sample from the second submission

If sampling from the first submission of signatures does not qualify a petition and a second submission of sufficient size is submitted before the deadline then a random sample of signatures is taken from the second submission for verification. This form of sampling is called stratified random sampling where the strata correspond to the two submissions of signatures comprising the petition. For denoting the various counts of signatures and electors the subscripts F and S will designate the first and second submissions, respectively.

Submissions:

N_F = number of signatures in first submission

N_S = number of signatures in second submission

Y_F = number of valid signatures in first submission

Y_S = number of valid signatures in second submission

D_F = number of duplicates of valid signatures from electors who contribute two or more valid signatures to the first submission

D_S = number of duplicates of valid signatures from electors who contribute no valid signatures to the first submission and two or more valid signatures to the second submission

D_{FS} = number of duplicates of valid signatures from electors who contribute one valid signature to the first submission and one or more valid signatures to the second submission

Samples:

n_F = size of combined sample from the first submission

n_S = size of sample from the second submission

y_F = number of valid signatures in the combined sample from first submission

y_S = number of valid signatures in the sample from second submission

Two valid signatures from electors

e_{F2} = number of electors with two valid signatures in the combined sample from first submission

e_{S2} = number of electors with no valid signature in the combined sample from first submission and two valid signatures in the sample from the second submission

e_{FS11} = number of electors with one valid signatures in the combined sample from first submission and one valid signature in the sample from the second submission

Three valid signatures from electors

e_{F3} = number of electors with three valid signatures in the combined sample from first submission

e_{S3} = number of electors with no valid signature in the combined sample from first submission and three valid signatures in the sample from the second submission

e_{FS12} = number of electors with one valid signature in the combined sample from first submission and two valid signatures in the sample from the second submission

e_{FS21} = number of electors with two valid signatures in the combined sample from first submission and one valid signature in the sample from the second submission

Note that all the counts of electors with duplicate valid signatures observed in the samples (e_{F2}, e_{S2}, e_{FS11}, e_{F3}, e_{S3}, e_{FS12}, and e_{FS21}) are for distinct electors except that the e_{FS21} electors with two signatures in the combined sample from the first submission and a third signature in the sample from the second submission are included in the count of electors with two signatures in the combined sample, e_{F2}. Since three signatures rarely occur in the samples, the counts involving three valid signatures (e_{F3}, e_{S3}, e_{FS12}, and e_{FS21}) are usually all zero.

The total numbers of signatures submitted and numbers of valid signatures for both submissions are found by adding the totals from each submission

$$N = N_F + N_S \quad \text{and} \quad Y = Y_F + Y_S \qquad \text{(Equation 12)}$$

The total number of duplicates of valid signatures in both submissions is partitioned into three components

$$D = D_F + D_S + D_{FS}, \qquad \text{(Equation 13)}$$

where the components correspond to which submissions contain the first two signatures in the petition from an elector: both in first, both in second, and one in each of the first and second submissions.

The sample size for the second submission, n_S, is taken as the larger of the two values: $n_F \left| \frac{N_S}{N_F} \right|$ and 250. The first value is proportional to the combined sample size from the first submission

$$n_S = n_F \left| \frac{N_S}{N_F} \right|, \qquad \text{(Equation 14)}$$

and correspondingly to using the same sampling rate as that for the first submissions, $n_S / N_S = n_F / N_F$. When the 5% sampling rate is adopted, Equation 14 applies for second submission sizes of 5,000 or larger and the minimum sample size n_S = 250 applies for second submission sizes less than 5,000.

For petitions that are submitted in two submissions, estimates for the total number of valid signatures and the total number of duplicates of valid signatures need to be calculated for substitution into Equation 1. The following steps describe the calculation of these estimates.

Steps for determining the result from the two submissions

1. Calculate the estimate for the number of valid signatures in the second submission by multiplying the number of valid signatures in the sample by the expansion factor calculated as the ratio of the submission size over the sample size

$$\hat{Y}_{\mathrm{S}} = \left(\frac{N_{\mathrm{S}}}{n_{\mathrm{S}}} \right) y_{\mathrm{S}} \qquad \text{(Equation 15)}$$

Add this estimate to the corresponding estimate for the number of valid signatures in the first submission $\hat{Y}_{\mathrm{F}}$, which was similarly calculated using Equation 9, to produce an estimate for the total number of valid signatures in the first and second submissions

$$\hat{Y} = \hat{Y}_{\mathrm{F}} + \hat{Y}_{\mathrm{S}} \qquad \text{(Equation 16)}$$

2. Calculate the estimate for the total number of duplicates, $\hat{D}$, by estimating the three components in Equation 13

$$\hat{D} = \hat{D}_{\mathrm{F}} + \hat{D}_{\mathrm{S}} + \hat{D}_{\mathrm{FS}}, \qquad \text{(Equation 17)}$$

where

$$\hat{D}_{\mathrm{F}} = \left(\frac{N_{\mathrm{F}}}{n_{\mathrm{F}}} \right)^2 \left(e_{\mathrm{F2}} + e_{\mathrm{F3}} \right) + \left(\frac{N_{\mathrm{F}}}{n_{\mathrm{F}}} \right) e_{\mathrm{F3}} + \left(\frac{N_{\mathrm{S}}}{n_{\mathrm{S}}} \right) e_{\mathrm{FS21}} \qquad \text{(Equation 18)}$$

$$\hat{D}_{\mathrm{S}} = \left(\frac{N_{\mathrm{S}}}{n_{\mathrm{S}}} \right)^2 \left(e_{\mathrm{S2}} + e_{\mathrm{S3}} \right) + \left(\frac{N_{\mathrm{S}}}{n_{\mathrm{S}}} \right) e_{\mathrm{S3}}, \qquad \text{(Equation 19)}$$

$$\hat{D}_{\mathrm{FS}} = \left(\frac{N_{\mathrm{F}}}{n_{\mathrm{F}}} \right) \left(\frac{N_{\mathrm{S}}}{n_{\mathrm{S}}} \right) \left(e_{\mathrm{FS11}} + e_{\mathrm{FS12}} \right) + \left(\frac{N_{\mathrm{S}}}{n_{\mathrm{S}}} \right) e_{\mathrm{FS12}}, \qquad \text{(Equation 20)}$$

The expansion factor in the first component of Equations 18, 19, and 20 are for the total numbers of electors with two or three valid signatures observed in the samples from both submission with the first two signatures from respectively: the first submission in Equation 18, the second submission in Equation 19, and the first signature from the first submission and the second signature from the second submission in Equation 20. The expansion factor in the other component(s), corresponding to three valid signatures, are included to cancel the contribution of the third valid signature to the estimate for the number of valid signatures in both submissions (Equations 16 and 21).

3. Substitute the estimates from steps 1 and 2 in Equation 1 to give the estimate for the total number of distinct valid signatures in the both submissions

$$\hat{M} = \hat{Y} - \hat{D} \qquad \text{(Equation 21)}$$

4. Compare the estimate $\hat{M}$ to the required number R. If $\hat{M}$ is greater than or equal to R, the petition qualifies for the ballot. If $\hat{M}$ is less than R, the petition does not qualify.

EXAMPLES

Calculations required for determining the qualification of petitions are made for three illustrative examples. The required numbers of valid signatures are based on percentages of the total votes for governor in the last general election. Based on the 2002 General Election: $R = 75{,}630$ (6%) and 100,840 (8%) for statutory and constitutional initiative petitions, respectively. *R* varies depending on the signature requirements for the type of petition submitted. This formula may also be applied to minor party, recall, and candidate petitions.

Example 1:
Qualification from the verification of the first sample

Suppose $N =$ 120,000 signatures are submitted in a statutory petition requiring $R = 75{,}630$ distinct valid signatures. The first sample size is fixed at $n_1 = 1{,}000$ and the second is taken as $n_2 = 5{,}000$ so that combined sample size $n = n_1 + n_2 = 6{,}000$ is equal to 5% of the petition size. From the first sample, suppose there are $y_1 = 770$ valid signatures. Follow steps 1-5 for determining the result from the first sample to calculate the following:

Quantity	Equation	Numerical result
Estimate for number valid	$\hat{Y}_1 = \left(\frac{N}{n_1}\right) y_1 = \left(\frac{120{,}000}{1{,}000}\right) 770 = (120)(770)$	$92{,}400.0$
Margin of error	$\text{MOE} = 1.645\sqrt{\left\{\frac{N(N-n_1)}{n_1 - 1}\right\}\left(\frac{y_1}{n_1}\right)\left(1-\frac{y_1}{n_1}\right)}$ $= 1.645\sqrt{\left\{\frac{120{,}000(120{,}000-1{,}000)}{1{,}000-1}\right\}\left(\frac{770}{1{,}000}\right)\left(1-\frac{770}{1{,}000}\right)}$	$2{,}617.3$
Lower limit for number valid	$Y_{\text{L}} = \hat{Y}_1 - \text{MOE} = 92{,}400 - 2{,}617.3$	$89{,}782.7$
Upper limit for # duplicates	$D_{\text{U}} = 0.08N = (0.08)(120{,}000)$	$9{,}600.0$
Lower limit for # distinct valid	$M_{\text{L}} = Y_{\text{L}} - D_{\text{U}} = 89{,}782.7 - 9{,}600$	$80{,}182.7$

Since $M_{\text{L}} = 80{,}182.7$ is greater than $R = 75{,}630$, the petition qualifies for the ballot.

Example 2:

Rejection from the combined first and second samples

Suppose $N = 125{,}000$ signatures are submitted in a constitutional petition requiring $R = 100{,}840$ distinct valid signatures. The first sample size is fixed at $n_1 = 1{,}000$ and the second taken as $n_2 = 5{,}250$ so that combined sample size $n = n_1 + n_2 = 6{,}250$ is equal to 5% of the petition size. From the first sample, suppose there are $y_1 = 820$ valid signatures. Follow steps 1-4 for determining the result from the first sample to calculate the following:

Quantity	Equation	Numerical result
Estimate for number valid	$\hat{Y}_1 = \left(\frac{N}{n_1}\right) y_1 = \left(\frac{125{,}000}{1{,}000}\right) 820 = (125)(820)$	$102{,}500.0$
Margin of error	$\text{MOE} = 1.645\sqrt{\left\{\frac{N(N-n_1)}{n_1 - 1}\right\}\left(\frac{y_1}{n_1}\right)\left(1 - \frac{y_1}{n_1}\right)}$ $= 1.645\sqrt{\left\{\frac{125{,}000(125{,}000 - 1{,}000)}{1{,}000 - 1}\right\}\left(\frac{820}{1{,}000}\right)\left(1 - \frac{820}{1{,}000}\right)}$	$2{,}489.4$
Lower limit for# valid	$Y_{\text{L}} = \hat{Y}_1 - \text{MOE} = 102{,}500 - 2{,}489.4$	$100{,}010.6$
Upper limit for # duplicates	$D_{\text{U}} = 0.08N = (0.08)(125{,}000)$	$10{,}000.0$
Lower limit for # distinct valid	$M_{\text{L}} = Y_{\text{L}} - D_{\text{U}} = 100{,}010.6 - 10{,}000$	$90{,}010.6$

Since $M_{\text{L}} = 90{,}010.6$ is less than $R = 100{,}840$, the second sample shall be verified.

From the second sample of $n_2 = 5{,}250$ signatures suppose there are $y_2 = 4{,}240$ valid signatures so that there are $y = y_1 + y_2 = 820 + 4{,}240 = 5{,}060$ valid signatures in the combined sample of size $n = 6{,}250$. Among the 5,060 valid signatures suppose that $e_2 = 14$ electors contribute two valid signatures and $e_3 = 1$ elector contributes three valid signatures. Follow the steps 1-4 for determining the result from the combined first and second samples.

Quantity estimated	Estimating equation	Numerical result
Number valid	$\hat{Y} = \left(\frac{N}{n}\right) y = \left(\frac{125,000}{6,250}\right) 5,060 = (20)(5,060)$	$101,200$
Number of duplicates	$\hat{D} = \left(\frac{N}{n}\right)^2 (e_2 + e_3) + \left(\frac{N}{n}\right) e_3$ $= \left(\frac{125,000}{6,250}\right)^2 (14+1) + \left(\frac{125,000}{6,250}\right) 1$ $= (400)(15) + (20)(1)$	$6,020$
Number of distinct valid	$\hat{M} = \hat{Y} - \hat{D} = 101,200 - 6,020$	$95,180$

Since $\hat{M} = 95,180$ is less than the required number $R = 100,840$, the petition does not qualify for the ballot.

Example 3:

Qualification from samples for two submissions.

Suppose now that the petition of size $N_{\mathrm{F}} = 125,000$ in Example 2 was only the first submission and that $N_{\mathrm{S}} = 20,000$ more signatures were submitted in the second submission. The 5% sampling rate was also applied to the second submission so the resulting sample size was $n_{\mathrm{S}} = 1,000$.

For the combined sample from the first submission the following were observed: $y_{\mathrm{F}} = 5,060$ valid signatures, $e_{\mathrm{F2}} =$ 14 electors with two valid signatures, and $e_{\mathrm{F3}} =$ 1 elector with three valid signatures. The number of valid signatures was estimated in Example 2 as $\hat{Y}_{\mathrm{F}} = 101,200$.

After the sample from the second submission was verified suppose the following counts were observed:

$y_{\mathrm{S}} = 790$ valid signatures in the sample from second submission

$e_{\mathrm{S2}} = 1$ elector with no valid signature in the combined sample from first submission and two valid signatures in the sample from the second submission

$e_{\mathrm{S3}} = 0$ electors with no valid signature in the combined sample from first submission and three valid signatures in the sample from the second submission

$e_{\mathrm{FS11}} = 2$ electors with one valid signature in the combined sample from first submission and one valid signature in the sample from the second submission

$e_{\mathrm{FS12}} = 0$ electors with one valid signatures in the combined sample from first submission and two valid signatures in the sample from the second submission

$e_{FS21} = 1$ elector with two valid signatures in the combined sample from first submission and one valid signature in the sample from the second submission

Follow the steps 1-3 for determining the result from both submissions for calculating the following estimates.

Quantity estimated	Estimating equation	Numerical result
# valid in 2nd submission	$\hat{Y}_S = \left(\frac{N_S}{n_S}\right) y_S = \left(\frac{20,000}{1,000}\right) 790$	$15,800$
# valid in both submissions	$\hat{Y} = \hat{Y}_F + \hat{Y}_S = 101,200 + 15,800$	$117,000$
# duplicates in three components	$\hat{D}_F = \left(\frac{N_F}{n_F}\right)^2 (e_{F2} + e_{F3}) + \left(\frac{N_F}{n_F}\right) e_{F3} + \left(\frac{N_S}{n_S}\right) e_{FS21}$ $= \left(\frac{125,000}{6,250}\right)^2 (14+1) + \left(\frac{125,000}{6,250}\right) 1 + \left(\frac{20,000}{1,000}\right) 1$ $= (400)(15) + (20)(1) + (20)(1)$	$6,040$
	$\hat{D}_S = \left(\frac{N_S}{n_S}\right)^2 (e_{S2} + e_{S3}) + \left(\frac{N_S}{n_S}\right) e_{S3}$ $= \left(\frac{20,000}{1,000}\right)^2 (1+0) + \left(\frac{20,000}{1,000}\right) 0$ $= (400)(1) + (20)(0)$	400
# duplicates in both submissions	$\hat{D}_{FS} = \left(\frac{N_F}{n_F}\right)\left(\frac{N_S}{n_S}\right)(e_{FS11} + e_{FS12}) + \left(\frac{N_S}{n_S}\right) e_{FS12}$ $= \left(\frac{125,000}{6,250}\right)\left(\frac{20,000}{1,000}\right)(2+0) + \left(\frac{20,000}{1,000}\right) 0$ $= (20)(20)(2) + (20)(0)$	800
	$\hat{D} = \hat{D}_F + \hat{D}_S + \hat{D}_{FS}$ =6,040+400+400	$7,240$
# distinct valid	$\hat{M} = \hat{Y} - \hat{D} = 117,000 - 6,840$	$109,760$

Since $\hat{M} = 109,760$ is greater than the required number $R = 100,840$, the petition qualifies for the ballot.