Estimating population abundance for simple random sampling

$\hat{X} = K\bar{X}$	the estimated population size. This is formula 8.6 from [?]
K	the number of sampling units in the landscape.
$\bar{X} = \frac{1}{k}\sum_{i=1}^{k} x_i$	the mean count per sampling unit.
k	the number of sampling units in the landscape.
x_i	the count in the i^{th} sampling unit.
$\hat{X} \pm t_{k-1,1-\frac{\alpha}{2}} SE(\hat{X})$	is the $100(1-\alpha)$ percent confidence interval. This is equation 8.13 from [?].
$t_{k-1,1-\frac{\alpha}{2}}$	the t-statistic with $k-1$ degrees of freedom, and where $\alpha = 0.05$ will evaluate the 95% confidence interval.
$SE(\hat{X}) = \sqrt{\frac{K^2}{k}\left(1-\frac{k}{K}\right)s^2}$	the standard error in the population size estimate. This is equation 8.8 in [?].
$s^2 = \frac{\sum_{i=1}^{k}(x_i-\bar{X})^2}{k-1}$	between-sampling unit variability.
$\hat{\sigma} = SE(\hat{X})\sqrt{k}$	estimated standard error.

Example

Let,

$$x_1 = 40 \qquad x_2 = 30 \qquad x_3 = 20 \qquad k = 3 \qquad \text{and} \qquad K = 10.$$

The estimated population abundance is,

$$\hat{X} = K\bar{X} = 10 \times \frac{1}{3}(40 + 30 + 20) = 10 \times 30 = 300.$$

We can use `R` or a t-table to calculate $t_{2,0.975}$ as,

```
>qt(0.975,2) = 4.30.
```

We estimate the between-sampling unit variability as,

$$s^2 = \frac{(40-30)^2 + (30-30)^2 + (20-30)^2}{3-1} = \frac{(10)^2 + (0)^2 + (-10)^2}{2} = 100,$$

and the standard error in the population estimate as,

$$SE(\hat{X}) = \sqrt{\frac{10^2}{3}\left(1 - \frac{3}{10}\right)100} = 10\sqrt{\frac{100}{3} \times 0.7} = 48.3.$$

As such, the lower bound on 95% confidence interval is,

$$300 - 4.3 \times 48.3 = 92,$$

and the upper bound on 95% confidence interval is,

$$300 + 4.3 \times 48.3 = 508.$$

A Q-Q plot can be used evaluate where the x_i's are normally distributed as assumed by the calculations.

Sample size calculation

$k^* = \left(\frac{Z_{1-\frac{\alpha}{2}}\ \hat{\sigma}}{\delta}\right)^2$	the minimum simple size needed to achieve a $100(1-\alpha)$ percent confidence interval with a half-width equal to δ. This is equation 2 from [?].
$Z_{1-\frac{\alpha}{2}}$	the value for which the cumulative density of the standard normal distribution is $1-\frac{\alpha}{2}$.
$\hat{\sigma} = SE(\hat{X})\sqrt{k}$	the estimated standard deviation of the population estimate.

We evaluate $Z_{1-\frac{\alpha}{2}}$ in `R` as `>qnorm(1-`$\frac{\alpha}{2}$`)`. Therefore, if $\alpha = 0.05$ we calculate k^* for a 95% confidence interval as,

`>qnorm(0.975) = 1.96`.

The estimated standard deviation is,

$$\hat{\sigma} = 48.3 \times \sqrt{3} = 83.7,$$

where $SE(\hat{X}) = 48.3$ and $k = 3$ are the values from the previous section.

If we want $\delta = 50$ then,

$$k^* = \left(\frac{1.96 \times 83.7}{50}\right)^2 = 10.7.$$

Therefore, since $K = 10$ and for a landscape that is as variable as this one ($\hat{\sigma} = 83.7$), we could only get a 95% confidence interval with a halfwidth of $\delta = 50$ if we sampled all the sampling units in the landscape.

References

[1] Gregoire, T.G., and D. L. R. Affleck. 2018. Estimating Desired Sample Size for Simple Random Sampling of a Skewed Population, The American Statistician, 72:2, 184-190, `DOI:10.1080/00031305.2017.1290548`.

[2] Skalski, J.R., Ryding, K.E., Millspaugh, J.J., Millspaugh, J., 2005. Wildlife Demography: Analysis of Sex, Age, and Count Data. Elsevier Science and Technology, Burlington, UNITED STATES. `https://ebookcentral-proquest-com.qe2a-proxy.mun.ca/lib/mun/detail.action?docID=269552`