Data Science 1
Statistical Inference
Estimation for One Population
Population Total

Ann Maharaj

Confidence Interval Estimation of the Population Total

- Given a sample of size n from a population of size N, and sample mean $\overline{X}$, a point estimator of the population total T is $\widehat{T} = N\overline{X}$.

- Like the sample mean, $\widehat{T}$ has a sampling distribution. The mean and standard error of this sampling distribution are:

$$E(\widehat{T}) = N \, x \, E(\overline{X})$$

$$SE(\widehat{T}) = N \, x \, SE(\overline{X})$$

- A confidence interval estimator of T is obtained, by forming a confidence interval for the mean μ and then multiplying the lower and upper limits by N.

Confidence Interval Estimation of the Population Total

- A $100(1-\alpha)\%$ confidence interval estimator of T when σ **is known** is

$$N\left(\overline{X} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$$

- The population size N is known but if $n/N > 0.05$ then including the *fpc*, the $100(1-\alpha)\%$ confidence interval estimator of T becomes

$$N\left(\overline{X} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}\right)$$

Confidence Interval Estimation of the Population Total

- A $100(1-\alpha)\%$ confidence interval estimator of T when **σ is unknown** is

$$N\left(\overline{X} \pm t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}\right)$$

- The population size N is known but if $n/N > 0.05$, then including the *fpc*, the $100(1-\alpha)\%$ confidence interval estimator of T becomes

$$N\left(\overline{X} \pm t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}\right)$$

Example

Consider again the taxi data in
Example 2 on Slide 12 in the set of slides "1Confidence intervals.pdf".

- The owner of a large fleet of taxis is trying to estimate his costs for next years operations.
- One major cost is for fuel purposes. Because of the high cost of petrol, and the need to reduce emissions, the owner has changed his entire fleet to hybrid vehicles.
- In order to estimate the fuel consumption of his hybrid fleet of taxis, he took a random sample of 8 taxis and measured the kilometres per litre achieved by each.
- The results are as follows: 3.5, 4.2, 5.3, 4.7, 3.4, 4.6, 4.9, 3.7. Obtain and report a 95% confidence interval estimate of the mean fuel consumption of all taxis in the fleet.
- Assume that the distribution of fuel consumption is approximately normal.

$n = 8$, $\overline{X}$ and s must be computed, $(1 - \alpha) = 0.95 \Rightarrow \alpha = 0.05 \Rightarrow \alpha/2 = 0.025$.

Example

- Here we determined from the formula

$$\overline{X} \pm t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}$$

  that the 95% confidence interval estimate of the mean fuel consumption of all taxis in the fleet is [3.7024, 4.8726] ≈ [3.70, 4.87] . Refer to *Slides 13-14 in the set of slides "1Confidence intervals.pdf"*.
- Suppose that the owner has 200 taxis in his fleet. Obtain the point estimate of the total, and obtain and report a 95% confidence interval estimate of the total consumption of the 200 taxis.
- We now require the formulae

$$\widehat{T} = N\overline{X}$$

$$N\left(\overline{X} \pm t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}\right)$$

Example

- To obtain the point estimate of the total we multiple the point estimate of the mean by the population size, $N = 200$. Hence, the point estimate of the total is 200 x 4.2875 = 857.50 kilometres per litre.

- To obtain the confidence interval estimate of the total, we multiple the lower limit and the upper limit by the population size. Hence, the 95% confidence interval estimate of the total consumption of the 200 taxis is
  $[200 \times 3.7024, 200 \times 4.8726] \approx [740.68, 974.52]$.

- We estimate the total consumption for all taxis in the fleet to be between approximately 740.68 and 974.52 kilometres per litre with 95% confidence.