Finite Population Correction

Numerical Statistics Fall, 2021

Jake Underland

2021-10-01

Contents

Lesson

Let $\Pi = \{x_1, x_2, \ldots, x_N\}$ be a finite population of size N, and suppose we want to estimate an unknown population mean.

The Population Mean

$$\mu = \frac{1}{N}\sum_{i=1}^{N} X_i$$

Sample Mean With Replacement

Draw a sample of size n from the above population ($\{X_1, X_2, \ldots, X_n\}$, $2 \leq n \leq N$). Then, the sample mean is

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

This is unbiased.

$$E(\bar{X}) = E(\frac{1}{n}\sum_{i=1}^{n} X_i) = \underbrace{\frac{1}{n}\sum_{i=1}^{n} E(X_i)}_{Linearity} = \frac{1}{n}\sum_{i=1}^{n} \mu = \mu$$

The above holds because under replacement, because

$$
\begin{aligned}
P(\underbrace{X_j}_{sample} = \underbrace{x_i}_{population}) &= \frac{1}{N} \quad \begin{cases} j = 1, 2, \ldots, n \\ i = 1, 2, \ldots, N \end{cases} \\
P(X_i = x_k, X_j = x_l) &= \frac{1}{N^2} \ldots \text{!!!} \\
\implies E(X_j) &= \sum_{i=1}^{N} x_i P(X_j = x_i) \\
&= \sum_{i=1}^{N} x_i \frac{1}{N} \\
&= \frac{1}{N} \sum_{i=1}^{N} X_i \\
&= \mu \ldots \text{by definition}
\end{aligned}
$$

However, we know this is unbiased regardless of whether you *sample with or without replacement*.

Sample Mean Without Replacement

In case of replacement,

$$
\begin{aligned}
P(X_j = x_i) &= \frac{1}{N} \quad \begin{cases} j = 1, 2, \ldots, n \\ i = 1, 2, \ldots, N \end{cases} \\
P(X_i = x_k, X_j = x_l) &= \begin{cases} \frac{1}{N}(\frac{1}{N-1}) & \ldots (k \neq l) \\ 0 & \ldots (k = l) \end{cases} \ldots \text{!!!} \\
\implies E(X_j) &= \sum_{i=1}^{N} x_i P(X_j = x_i) \\
&= \sum_{i=1}^{N} x_i \frac{1}{N} \\
&= \frac{1}{N} \sum_{i=1}^{N} X_i \\
&= \mu \ldots \text{by definition}
\end{aligned}
$$

Because while the joint marginal probability changes from the case with replacement, the marginal probability does not, meaning the same proof holds for unbiasedness of the sample mean in the case of no replacement. Why does the marginal probability not change? Think of this case ↓

$$
\begin{aligned}
P(X_3 = a) &= P(X_1 \neq a \cap X_2 \neq a \cap X_3 = a) \\
&= \underbrace{\frac{N-1}{N}}_{\text{remove } a} \cdot \underbrace{\frac{N-2}{N-1}}_{\text{remove } a \text{ and } X_1} \cdot \underbrace{\frac{1}{N-2}}_{\text{remove } X_1, X_2} \\
&= \frac{1}{N}
\end{aligned}
$$

The Population Variance

$$
\sigma_2 = \frac{1}{N} \sum_{i=1}^{N} (x_i - \mu)^2
$$

The Sample Variance With Replacement

$$Var(X_j) = \sigma^2$$

Proof:

$$\begin{aligned}
Var(X_j) &= E[(X_j - E(X_j))^2] \\
&= E[(X_j - \mu)^2] \\
&= \sum_{i=1}^{N} (x_i - \mu)^2 P(X_j = x_i) \\
&= \sum_{i=1}^{N} (x_i - \mu)^2 \frac{1}{N} \\
&= \frac{1}{N} \sum_{i=1}^{N} (x_i - \mu)^2 \\
&= \sigma^2 \ \square
\end{aligned}$$

Variance of Sample Mean With Replacement

$$Var(\bar{X}) = \frac{\sigma^2}{n}$$

Proof.

$$\begin{aligned}
Var(\bar{X}) &= Var(\frac{1}{n} \sum_{i=1}^{n} X_i) \\
&= \frac{1}{n^2} \sum_{i=1}^{n} Var(X_i) \ldots \textit{Independence} \\
&= \frac{1}{n^2} \sum_{i=1}^{n} \sigma^2 \\
&= \frac{\sigma^2}{n} \ldots \square
\end{aligned}$$

Variance of Sample Mean Without Replacement

First, we start with covariance.

$$Cov(X_i, X_j) = -\frac{\sigma^2}{N-1} \ \ (i \neq j)$$

Proof.

$$\begin{aligned}
Cov(X_i, X_j) &= E[(X_i - E[X_i])(X_j - E[X_j])] \\
&= E[(X_i - \mu)(X_j - \mu)] \\
&= \sum_{i=1, j=1}^{N} (X_i - \mu)(X_j - \mu) \\
&= \sum_{k=l, k \neq l}^{N} (x_k - \mu)(x_l - \mu) \underbrace{P(X_i = x_k, X_j = x_l)}_{= 0 \text{ when } k = l} \\
&= \sum_{k \neq l}^{N} (x_k - \mu)(x_l - \mu) \frac{1}{N} (\frac{1}{N-1}) \\
&= \frac{1}{N(N-1)} \sum_{k \neq l}^{N} (x_k - \mu)(x_l - \mu)
\end{aligned}$$

We know that

$$\{\sum_{i=1}^{N}(x_i-\mu)\}^2 = 0$$

$$\implies \sum_{i=1}^{N}(x_i-\mu)^2 + \sum_{k\neq l}(x_k-\mu)(x_l-\mu) = 0$$

$$\begin{aligned}\implies \sum_{k\neq l}(x_k-\mu)(x_l-\mu) &= -\sum_{i=1}^{N}(x_i-\mu)^2 \\ &= -N\sigma^2\end{aligned}$$

Thus,

$$\begin{aligned}Cov(X_i,X_j) &= \frac{1}{N(N-1)}\sum_{k\neq l}^{N}(x_k-\mu)(x_l-\mu) \\ &= \frac{1}{N(N-1)}(-N\sigma^2) \\ &= -\frac{\sigma^2}{N-1}\ldots\square\end{aligned}$$

Using this, we find that

$$Var(\bar{X}) = \frac{\sigma^2}{n}(\frac{N-n}{N-1})$$

Proof.

$$\begin{aligned}Var(\bar{X}) &= Var(\frac{1}{n}\sum_{i=1}^{n}X_i) \\ &= \frac{1}{n^2}Var(\sum_{i=1}^{n}X_i) \\ &= \frac{1}{n^2}\sum_{i=1}^{n}Var(X_i) + \frac{1}{n^2}\sum_{i\neq j}Cov(X_i,X_j) \\ &= \frac{\sigma^2}{n} + \frac{n-1}{n}(-\frac{\sigma^2}{N-1}) \\ &= \frac{\sigma^2}{n}(1-\frac{n-1}{N-1}) \\ &= \frac{\sigma^2}{n}\underbrace{(\frac{N-n}{N-1})}_{\text{...Finite Population Correction Factor}} \qquad \square\end{aligned}$$

Notes on the Finite Population Correction Factor

We call the constant $FPC = \sqrt{\frac{N-n}{N-1}}$ the finite population correction factor.
We can use the FPC to construct confidence intervals for means or proportions when sampling from a finite population without replacement.
The confidence interval for μ is given by:

$$\bar{X} \pm 1.96\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

Proof.

$$\frac{\bar{X} - E(\bar{X})}{\sqrt{Var(\bar{X})}} \sim N(0,1)$$

$$\implies P(-1.96 < \frac{\bar{X} - E(\bar{X})}{\sqrt{Var(\bar{X})}} < 1.96) = 0.95$$

$$\implies P(-1.96 < \frac{\bar{X} - \mu}{\sqrt{\frac{\sigma^2}{n}(\frac{N-n}{N-1})}} < 1.96) = 0.95$$

$$\implies P(\bar{X} - 1.96\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} < \mu < \bar{X} + 1.96\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}) = 0.95 \ldots \square$$

When sampling is with replacement, $Var_{(1)}(\bar{X}) = \frac{\sigma^2}{n}$. When without replacement, $Var_{(2)}(\bar{X}) = \frac{\sigma^2}{n}(\frac{N-n}{N-1})$.

$$Var_{(1)}(\bar{X}) = \frac{\sigma^2}{n} > \frac{\sigma^2}{n}(\frac{N-n}{N-1}) = Var_{(2)}(\bar{X}) \ldots \text{because } (n > 1)$$

From here we see that using the sample mean to estimate μ when drawing from a finite population without replacement has higher accuracy than doing so with replacement. However, the difference is negligible the higher the cardinality of population N and the lower the sample size n.

Exercises

Exercise 01

Solve a linear equation $\sum_{i=1}^{n}(x_i - c) = 0$ for c, and verify that its solution is the arithmetic mean $\bar{x}$.

Solution.

$$\sum_{i=1}^{n}(x_i - c) = 0$$
$$\implies \sum_{i=1}^{n} x_i = nc$$
$$\implies c = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}$$