Causal Inference

Justin Grimmer

Associate Professor
Department of Political Science
University of Chicago

March 30th, 2018

Population

Definition (Population)

Consider a population with N units that have fixed values of an attribute $X_1, X_2, ..., X_N$.

X_i is the fixed value of the attribute for the i-th member of the population.

The population mean is given by

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

and the population variance is given by

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

Simple Random Sampling

Definition (Random Sample with replacement)

Assume a sample of n elements that are drawn with replacement and the attribute values of the sample members are denoted by $x_1, x_2, x_3, ..., x_n$. Each x_i is a random variable, i.e. the value of the i-th member of the sample. Sampling with replacement is equivalent to iid sampling. The draws $x_1, ..., x_n$ are independent and identically distributed random variables.

Definition (Random Sample without replacement)

Assume a sample of n elements that are drawn without replacement and the attribute values of the sample members are denoted by $x_1, x_2, x_3, ..., x_n$. Each x_i is a random variable, i.e. the value of the i-th member of the sample.

- For sampling without replacement, each of the $\binom{N}{n}$ possible samples of size n taken without replacement has the same probability of occurrence
- Sampling without replacement is fairly similar to the iid sampling with replacement, except that the draws $x_1, ..., x_n$ are no longer independent.
    - The covariance between two draws x_i and x_j is $Cov[x_i, x_j] = -\frac{\sigma^2}{(N-1)}$ if $i \neq j$
- Formulas for sampling with replacement still offer a good approximation as long as the sample size n is small relative to the population size N

Expected Value and Variance of each Sample Member

Lemma (Expected Value and Variance of each Sample Member)

Write distinct attribute values of the population members by $\theta_1, \theta_2, ..., \theta_m$ *and denote the number of population members that have the value* θ_j *by* n_j*. Then* x_i *is a discrete random variable with a simple PMF which is*

$$P(x_i = \theta_j) = \frac{n_j}{N} \text{ with } E[x_i] = \mu \text{ and } Var[x_i] = \sigma^2$$

This holds for sampling with and without replacement.

Expected Value and Variance of each Sample Member

Proof.

x_i can only take on possible values $\theta_1, \theta_2, \ldots, \theta_m$ and since each population member is equally likely to be the i-th member in the sample it follows that the probability that x_i assumes the value θ_j is $\frac{n_j}{N}$. The expected value then is

$$
\begin{aligned}
E[x_i] &= \sum_{j=1}^{m} \theta_j P(x_i = \theta_j) && (1.1)\\
&= \sum_{j=1}^{m} \theta_j \frac{n_j}{N} && (1.2)\\
&= \frac{1}{N} \sum_{j=1}^{m} n_j \theta_j && (1.3)\\
&= \frac{1}{N} N\mu && (1.4)\\
&= \mu && (1.5)\\
& && (1.6)
\end{aligned}
$$

the second to last equality follows because there are n_j population members with value θ_j and therefore the sum is equal to the sum of the values of all population members.

□

Expected Value and Variance of each Sample Member

Proof.

The variance of each sample member is

$$
\begin{aligned}
Var[x_i] &= E[x_i^2] - [E[x_i]]^2 && (1.7)\\
&= \sum_{j=1}^{m} \theta_j^2 P(x_i = \theta_j) - \mu^2 && (1.8)\\
&= \frac{1}{N}\sum_{j=1}^{m} n_j \theta_j^2 - \mu^2 && (1.9)\\
&= \frac{1}{N}\sum_{i}^{N} X_i^2 - \mu^2 && (1.10)\\
&= \frac{1}{N}(\sum_{i}^{N} X_i^2 - 2N\mu^2 + N\mu^2) && (1.11)\\
&= \frac{1}{N}(\sum_{i}^{N} X_i^2 - 2\mu\sum_{i}^{N} X_i + \sum_{i}^{N}\mu^2) && (1.12)\\
&= \frac{1}{N}\sum_{i}^{N}(X_i - \mu)^2 && (1.13)\\
&= \sigma^2 && (1.14)\\
& && (1.15)
\end{aligned}
$$

□

Sample Mean Estimator

Theorem (Sample Mean Unbiased for Population Mean)

Let $\bar{x}$ be the sample mean estimator given by $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$. Then for simple random sampling we have that the sample mean is an unbiased estimator of the population mean

$$E[\bar{x}] = \mu$$

Proof.

Using the lemma $E[x_i] = \mu$ we have

$$E[\bar{x}] = \frac{1}{n}E[\sum_{i=1}^{n} x_i] = \frac{1}{n} n\,\mu = \mu$$

This holds for sampling with and without replacement. □

Variance of the Sample Mean

Theorem (Variance of the Sample Mean)

For simple random sampling with replacement the variance of the sample mean is

$$Var[\bar{x}] = \frac{\sigma^2}{n}$$

For simple random sampling without replacement the variance of the sample mean is

$$Var[\bar{x}] = \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right) \quad (1.16)$$

$$= \frac{\sigma^2}{n}\left(1 - \frac{n-1}{N-1}\right) \quad (1.17)$$

where $\left(1 - \frac{n-1}{N-1}\right)$ *is often called the finite population correction.*

If the sampling fraction $\frac{n}{N}$ *is small, the finite population correction is close to one and therefore the variance with replacement is a good approximation. If n grows close to N the variance goes to zero.*

Variance of the Sample Mean

Proof.

Recall that for a linear combination of random variables the following result holds:

$$Var[a + \sum_{i=1}^{n} b_i x_i] = \sum_{i=1}^{n} \sum_{j=1}^{n} b_i b_j Cov[x_i, x_j]$$

where a and b are constants. Since the sample mean is a linear combination $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$ we have that

$$Var[\bar{x}_i] = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} Cov[x_i, x_j]$$

If we sample with replacement the x_i draws are independent and therefore $Cov[x_i, x_j] = 0$, but $Cov[x_i, x_i] = Var[x_i] = \sigma^2$ and therefore we would get the usual variance

$$Var[\bar{x}_i] = \frac{1}{n^2} \sum_{i=1}^{n} Var[x_i] \quad (1.18)$$

$$= \frac{\sigma^2}{n} \quad (1.19)$$

□

Variance of the Sample Mean

Proof.

Sampling without replacement induces a dependence among the x_i such that the covariance between two draws is given by

$$Cov[x_i, x_j] = -\frac{\sigma^2}{(N-1)} \text{ if } i \neq j$$

Using this result we can derive the variance for the sample mean for sampling without replacement as

$$\begin{aligned}
Var[\bar{x}_i] &= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} Cov[x_i, x_j] && (1.20)\\
&= \frac{1}{n^2}\sum_{i=1}^{n} Var[x_i] + \frac{1}{n^2}\sum_{i=1}^{n}\sum_{i\neq j} Cov[x_i, x_j] && (1.21)\\
&= \frac{\sigma^2}{n} - \frac{1}{n^2} n(n-1)\frac{\sigma^2}{N-1} && (1.22)\\
&= \frac{\sigma^2}{n}\left(1 - \frac{n-1}{N-1}\right) && (1.23)
\end{aligned}$$

□

Unbiased Estimator for Population Variance

Theorem (Unbiased Estimator for Population Variance)

Recall that the population variance is given by

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

Typically we use a sample analog estimator and estimate the population variance by the simple sample variance

$$\hat{\sigma}_n^2 = \frac{1}{n}\sum_{i}^{n}(x_i - \bar{x})^2$$

For sampling with replacement this estimator is biased since $E[\hat{\sigma}_n^2] = \frac{n-1}{n}\sigma^2$.

So an unbiased estimator can be obtained by multiplying $\hat{\sigma}_n^2$ *with a factor of* $\frac{n}{n-1}$ *(often called the Bessel correction) which leads to the commonly used modified sample variance estimator*

$$\hat{\sigma}_{n-1}^2 = \frac{1}{n-1}\sum_{i}^{n}(x_i - \bar{x})^2$$

which is unbiased so that $E[\hat{\sigma}_{n-1}^2] = \sigma^2$.

Unbiased Estimator for Population Variance

Proof.

$$
\begin{aligned}
E[\sigma_n^2] &= E\left[\frac{1}{n}\sum_{i=1}^{n}\left(x_i - \frac{1}{n}\sum_{j=1}^{n} x_j\right)^2\right] && (1.24)\\
&= \frac{1}{n}\sum_{i=1}^{n} E\left[x_i^2 - \frac{2}{n}x_i\sum_{j=1}^{n}x_j + \frac{1}{n^2}\sum_{j=1}^{n}x_j\sum_{k=1}^{n}x_k\right] && (1.25)\\
&= \frac{1}{n}\sum_{i=1}^{n}\left[\frac{n-2}{n}E[x_i^2] - \frac{2}{n}\sum_{j\neq i}E[x_i x_j] + \frac{1}{n^2}\sum_{j=1}^{n}\sum_{k\neq j}^{n}E[x_j x_k] + \frac{1}{n^2}\sum_{j=1}^{n}E[x_j^2]\right] && (1.26)\\
&= \frac{1}{n}\sum_{i=1}^{n}\left[\frac{n-2}{n}(\sigma^2+\mu^2) - \frac{2}{n}(n-1)\mu^2 + \frac{1}{n^2}n(n-1)\mu^2 + \frac{1}{n}(\sigma^2+\mu^2)\right] && (1.27)\\
&= \frac{n-1}{n}\sigma^2. && (1.28)\\
& && (1.29)
\end{aligned}
$$

□

Unbiased Estimator for the Variance of the Sample Mean

Theorem (Unbiased Estimator for the Variance of the Sample Mean)

Recall that for sampling with replacement the variance of the sample mean is given by

$$Var[\bar{x}_i] = \frac{\sigma^2}{n}$$

And the modified sample variance

$$\hat{\sigma}^2_{n-1} = \frac{1}{n-1}\sum_{i}^{n}(x_i - \bar{x})^2$$

is an unbiased estimator for the population variance so that $E[\hat{\sigma}^2_{n-1}] = \sigma^2$.

So by plugging in we obtain an unbiased estimator for the variance of the sample mean

$$Var[\bar{x}] = \frac{\hat{\sigma}^2_{n-1}}{n}$$

Unbiased Estimator for Population Variance

Theorem (Unbiased Estimator for Population Variance)

For simple random sampling without replacement, the population variance is the same

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

but it turns out that

$$E[\hat{\sigma}_n^2] = \sigma^2 \left(\frac{n-1}{n}\right)\frac{N}{N-1}$$

so the simple sample variance estimator $\hat{\sigma}_n^2$ *is also biased for the population variance. However, an unbiased estimator can be obtained by multiplying* $\hat{\sigma}_n^2$ *by* $\frac{n}{n-1}\frac{N-1}{N}$.

Unbiased Estimator for Population Variance

Proof.

$$\hat{\sigma}_n^2 = \frac{1}{n}\sum_{i}^{n}(x_i - \bar{x})^2 = \frac{1}{n}\sum_{i}^{n} x_i^2 - \bar{x}^2 \tag{1.30}$$

$$E[\hat{\sigma}_n^2] = \frac{1}{n}\sum_{i=1}^{n} E[x_i^2] - E[\bar{x}^2] \tag{1.31}$$

Now we know that

$$E[x_i^2] = Var[x_i] + [E[x_i]]^2 = \sigma^2 + \mu^2 \tag{1.32}$$

$$E[\bar{x}^2] = Var[\bar{x}] + [E[\bar{x}]]^2 = \frac{\sigma^2}{n}\left(1 - \frac{n-1}{N-1}\right) + \mu^2 \tag{1.33}$$

so substituting in these expressions we get

$$E[\hat{\sigma}_n^2] = \frac{1}{n}\sum_{i=1}^{n} E[x_i^2] - E[\bar{x}^2] \tag{1.34}$$

$$= \frac{1}{n}\sum_{i=1}^{n} \sigma^2 + \mu^2 - \left(\frac{\sigma^2}{n}\left(1 - \frac{n-1}{N-1}\right) + \mu^2\right) \tag{1.35}$$

$$= \sigma^2\left(\frac{n-1}{n}\right)\frac{N}{N-1} \tag{1.36}$$

Unbiased Estimator for the Variance of the Sample Mean

Theorem (Unbiased Estimator for the Variance of the Sample Mean)

Recall that for sampling without replacement the variance of the sample mean is given by

$$Var[\bar{x}] = \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)$$

So by plugging in the unbiased estimator of the population variance, we obtain an unbiased estimator of the variance of the sample mean is given by

$$\hat{\sigma}^2_{\bar{x}} = \frac{\hat{\sigma}_n}{n}\left(\frac{N-n}{N-1}\right)\frac{n}{n-1}\frac{N-1}{N} \quad (1.37)$$

$$= \frac{\hat{\sigma}^2_{n-1}}{n}\left(1-\frac{n}{N}\right) \quad (1.38)$$

where $\hat{\sigma}^2_{n-1} = \frac{1}{n-1}\sum_i^n (x_i - \bar{x})^2$. So the variance estimator is the same as under sampling with replacement but multiplies it with the finite population correction $\left(1-\frac{n}{N}\right)$.

Types of Randomization

1. Simple random assignment
    - For each unit i we flip a coin, ie. randomly sample from a Bernoulli distribution where the probability of success is p
    - Equivalent to random sampling with replacement: the treatment assignment for unit, D_i, is a random variable with $P(D_i = 1) = p$ so each unit has the same probability of receiving the treatment
    - The vector of treatment assignments for all the units $D_1, D_2, ..., D_N$ constitute an iid sample
2. Complete random assignment
    - N_1 of N units are randomly assigned to treatment and N_0 units to control (with $N = N_1 + N_0$)
    - Equivalent to random sampling without replacement: each unit has the same probability of receiving the treatment, $p = N_1/N$, because each vector of treatment assignments $\binom{N}{N_1}$ is equally likely
    - The treatment assignments for the units $D_1, D_2, ..., D_N$ are not independent

Randomization for Causal Inference

- Random assignment allows for unbiased estimation of missing counterfactuals, such the unobserved average outcome under treatment $E[Y_{1i}]$, just like random sampling allows for unbiased estimation of the population mean.
- All units have a potential outcome under treatment Y_{1i} and we need to find the unobserved average outcome under treatment $E[Y_{1i}]$
- For the units that are assigned to the treatment we observe $Y_{1i} = Y_i$
- So if the units are randomly assigned to the treatment, this is like randomly sampling some Y_{1i}s from the population of all Y_{1i}s
- So the average observed outcome of the randomly chosen treated units will be an unbiased and consistent estimator of the average outcome under treatment in the population of all units $E[Y_{1i}]$
- This will allows us to identify the average treatment effects $\tau_{ATE} = E[Y_{1i}] - E[Y_{0i}]$