Population and Sampling Framework and Notation for One Factor Study

Introduction

This document summaries notation for units in populations and samples for a one factor study. We follow the convention of using upper case Latin letters for random variables, and Greek letters for parameters. We use lower case Latin letters for realized values of random variables. Although the letters and characters used here are arbitrary, when describing a 1 factor study, it will be easiest to communicate clearly if the same notation is used in this course.

Enumeration of Factors

$i=1,...,A$

Enumeration of Elements in the Population

$p=1,..., M$

Parameters in the Population when All Elements are Given Factor Level i:

μ_{ip} = response for element "p" when given Factor Level i
(Note: There is no response error so that $y_{ip}=\mu_{ip}$.)

Mean Response for Population Given Factor Level i

$$\mu_i=\sum_{p=1}^{M}\frac{\mu_{ip}}{M}.$$

Average of Population Means over all Factor Levels

$$\mu=\sum_{i=1}^{A}\frac{\mu_i}{A}$$

Effect of the Factor Level i: $\alpha_i = \mu_i - \mu$

Note that by definition,

$$\sum_{i=1}^{A}\alpha_i=0$$

Variance of Population Elements when Given Factor Level i:

$$\sum_{p=1}^{M}\frac{(\mu_{ip}-\mu_i)^2}{M}=\sigma_i^2\left(\frac{M-1}{M}\right) \quad where \quad \sigma_i^2=\sum_{p=1}^{M}\frac{(\mu_{ip}-\mu_i)^2}{M-1}$$

Sampling Elements From the Population:

In many setting, sampling of elements from the population consists of simple random sampling with replacement, or simple random sampling without replacement. Experimental studies often are described as having randomizing elements to A factor level groups. This process is equivalent to selecting A distinct simple random samples of elements without replacement from the population, and then assigning one sample to each level of the factor. Alternatively, we can think of A potentially observable populations of elements (with each population corresponding to M elements with a specific factor level), where simple random samples without replacement are selected from each population such that no two elements are included in more than one sample. The two descriptions describe the same problem. We define notation to use in this setting.

Order of selection of elements in a sample of size n_i .

$j=1,...,n_i$

Random variable representing response for jth selected element from the population of elements given Factor Level i, and deviation of that response from the population mean.

$Y_{ij} = \mu_i + E_{ij}$

or

$Y_{ij} = \mu + \alpha_i + E_{ij}$

Realized value for jth selected element from the population of elements given Factor Level i. The realized values are fixed constants.

$y_{ij} = \mu_i + e_{ij}$

Note that if the jth selected element is element p, then $(y_{ij}|j=p) = \mu_{ip.}$

Expected Values and Variances

Expected value of response for the jth selected element from the population given factor level i.

$E_P[Y_{ij}]=\mu_i$ Also $E_P[E_{ij}]=0$

Variance of response for the jth selected element from the population given factor level i.

$Var_P[Y_{ij}] = [(M-1)/M]\ \sigma_i^2$ Also, $Var_P[E_{ij}] = [(M-1)/M]\ \sigma_i^2$