LECTURE 17 PART II: **13.1 Effect of Nonresponse on Accuracy of Estimates**

Let N = population size,
N_1=total number of potential responding population elements,
N_2=total number of potential nonresponding population elements,
$N = N_1 + N_2$,
$\bar{X}_1 = X_1/N_1$=mean response for potential responding population elements,
$\bar{X}_2 = X_2/N_2$=mean response for potential nonresponding population elements,
$\bar{X} = (X_1 + X_2)/N = (N_1\bar{X}_1 + N_2\bar{X}_2)/N$=mean population response for all elements.

Suppose we sample n elements, of which n_1 are responding elements. Suppose also we don't get any response from the $n - n_1$ nonresponding elements. Then the sample mean $\bar{x}$ is an unbiased estimator of $\bar{X}_1$, that is, $\mathbb{E}(\bar{x}) = \bar{X}_1$, and the bias of $\bar{x}$ is

$$bias(\bar{x}) = \mathbb{E}(\bar{x}) - \bar{X} = \bar{X}_1 - \bar{X} = \bar{X}_1 - \frac{N_1\bar{X}_1 + N_2\bar{X}_2}{N}$$

$$= \Big(1 - \frac{N_1}{N}\Big)\bar{X}_1 - \frac{N_2}{N}\bar{X}_2 = \frac{N_2}{N}\big(\bar{X}_1 - \bar{X}_2\big).$$

Note that the bias doesn't depend on n_1. Hence, the bias cannot be reduced by drawing a larger sample (that is, increasing n_1). To reduce the bias, we have to reduce the ratio N_2/N= proportion of potential nonrespondents.