Cochran, W (1977). Sampling Techniques, 3rd ed. New York: Wiley.

ɔn contribute nothing to the
y of stores to estimate total
certain area sampling units
ɔossible, by expenditure of
ɪte nothing, so that in our

$V(\hat{Y}_j)$ is reduced when N_j is

estimators analog

the domain of interest (i.e.,

$$-\frac{N_j}{N}\Big) \qquad (2.60)$$

gligible,

$$(2.61)$$

$$-\frac{n}{N}\Big) \qquad (2.62)$$

of size n_j from them. The

$$\left(1-\frac{n_j}{N_j}\right) \qquad (2.63)$$

.62) the average number of
$_j = nP_j$ in (2.63), so that the
with both methods, (2.63)

$$(2.64)$$

$$\frac{C_j^2}{C_j^2+Q_j} \qquad (2.65)$$

the nonzeros. As might be
e of N_j is greater when the
relatively little among the
ee Jessen and Houseman

2.14 COMPARISONS BETWEEN DOMAIN MEANS

Let $\bar{y}_j$, $\bar{y}_k$ be the sample means in the jth and kth of a set of domains into which the units in a simple random sample are classified. The variance of their difference is

$$V(\bar{y}_j - \bar{y}_k) = V(\bar{y}_j) + V(\bar{y}_k) \qquad (2.66)$$

This formula applies also to the difference between two ratios $\hat{R}_j$ and $\hat{R}_k$.

One point should be noted. It is seldom of scientific interest to ask whether $\bar{Y}_j = \bar{Y}_k$, because these means would not be exactly equal in a finite population, except by a rare chance, even if the data in both domains were drawn at random from the same infinite population. Instead, we test the null hypothesis that the two domains were drawn from *infinite* populations having the same mean. Consequently we omit the fpc when computing $V(\bar{y}_j)$ and $V(\bar{y}_k)$, using the formula

$$V(\bar{y}_j - \bar{y}_k) = \frac{S_j^2}{n_j} + \frac{S_k^2}{n_k} \qquad (2.67)$$

A formula similar to (2.67) is obtained for tests of significance if one frames the question: Could the samples from the two domains have been drawn at random from the same *finite* population?

Under this null hypothesis it may be proved (see exercise 2.16) that

$$V(\bar{y}_j - \bar{y}_k) = S_{jk}^2\left(\frac{1}{n_j} + \frac{1}{n_k}\right)$$

where S_{jk}^2 is the variance of the finite population consisting of the combined domains.

2.15 VALIDITY OF THE NORMAL APPROXIMATION

Confidence that the normal approximation is adequate in most practical situations comes from a variety of sources. In the theory of probability much study has been made of the distribution of means of random samples. It has been proved that for any population that has a finite standard deviation the distribution of the sample mean tends to normality as n increases (see, e.g., Feller, 1957). This work relates to infinite populations.

For sampling without replacement from finite populations, Hájek (1960) has given necessary and sufficient conditions under which the distribution of the sample mean tends to normality, following work by Erdös and Rényi (1959) and Madow (1948). Hájek assumes a sequence of values n_ν, N_ν tending to infinity in such a way that $(N_\nu - n_\nu)$ also tends to infinity. The measurements in the νth population are denoted by $y_{\nu i}$, $(i = 1, 2, \ldots, N_\nu)$. For this population, let $S_{\nu\tau}$ be the set of units in the population for which

$$|y_{\nu i} - \bar{Y}_\nu| > \tau\sqrt{n_\nu(1-f_\nu)}S_\nu$$