Note on Adapting Davidson-Duclos Dominance Tests to Account for Sampling Designs of Household Surveys

(1) Davidson & Duclos (1998) show that in the case of simple random sampling, where n observations are distributed *iid*, if we define

$$D^s(z) = \tfrac{1}{n(s-1)!}\sum_{i=1}^{n}(z-y_i)^{s-1} I(y_i \le z)$$

which is equivalent to $\frac{z}{a!}P_a$ where $s = a+1$, then the estimator for the variance of $D^s(z)$ is simply

$$\hat{V} = \tfrac{1}{n}\sum_{i=1}^{n}\left[\tfrac{1}{a!}(z-y_i)^a \, I(y_i \le z)\right]^2 - \left(D^s(z)\right)^2$$

since $D^s(z)$ is the sum of *iid* random variables. It is easy to show that

$$\hat{V} = \left(\frac{z}{a!}\right)^2 \left[\frac{1}{n}\sum_{i=1}^{n}\left[\left(\frac{z-y_i}{z}\right)^a I(y_i \le z)\right]^2 - (P_a)^2\right] = \left(\frac{z}{a!}\right)^2 M$$

Simple t-statistics can now be used to test the difference between FGT measures for independent distributions A and B,

$$t = \frac{D_A^s - D_B^s}{\sqrt{\operatorname{var}(D_A^s) + \operatorname{var}(D_B^s)}}$$

$$= \frac{\frac{z}{a!}\left(P_a^A - P_a^B\right)}{\frac{z}{a!}\sqrt{M_A + M_B}} = \frac{P_a^A - P_a^B}{\sqrt{M_A + M_B}}$$

This is the basis for dominance testing between two simple randomly drawn distributions where each observation has equal probability of being in the sample.

(2) Before extending this to complex survey designs, let's introduce sampling weights (expansion factors) and simplify notation. Define the sampling weight as w_i, such that the sum of the weights is an estimate of the population size (or the population number of observations),

$$\hat{N} = \sum_{i=1}^{n} w_i \ .$$

Now for each value of s and z (i.e. for each test point in the dominance tests) define the following

$$x_i = \tfrac{1}{(s-1)!}(z - y_i)^{s-1} I(y_i \le z).$$

It now follows that

$$D^s(z) = \frac{1}{\hat{N}} \sum_{i=1}^{n} w_i x_i$$

and

$$\hat{V} = \frac{1}{\hat{N}} \sum_{i=1}^{n} w_i x_i^2 - \left(D^s(z)\right)^2 .$$

As a reminder, this estimator for the variance of $D^s(z)$ follows from the independence of each of the observations in the sample. While the estimate of $D^s(z)$ will not be affected by complex sampling designs (provided that our estimate uses the sampling weights), the variance estimator will. We will now see how to adjust our estimator for the variance of $D^s(z)$ (and tangentially, $D^s(z)$ itself) to acquire corrected test statistics for our dominance tests.

(3) Let's introduce some notation first:

<u>Population</u>:

Strata:	$h = 1, \ldots, L$
PSU$_i$ in strata h:	$i = 1, \ldots, M_h$
Elements in PSU$_{hi}$:	$j = 1, \ldots, N_{hi}$

Note: $N = \sum_{h=1}^{L} \sum_{i=1}^{M_h} N_{hi}$ = Total number of elements in the ***population***

Define: X_{hij} = modified poverty gap of household j in PSU i in strata h (as defined in part 2, above)

Then the population total is $X = \sum_{h=1}^{L} \sum_{i=1}^{M_h} \sum_{j=1}^{N_{hi}} X_{hij}$

<u>Sample</u>:

Strata:	$h = 1, \ldots, L$	(unchanged)
PSU$_i$ in strata h:	$i = 1, \ldots, m_h$	(sample of PSUs in strata h)
Elements in PSU$_{hi}$:	$j = 1, \ldots, n_{hi}$	(sample of elements in PSU$_{hi}$)
Sampling weight:	w_{hij}	

Note: $n = \sum_{h=1}^{L} \sum_{i=1}^{m_h} n_{hi}$ = Total number of elements in the ***sample***

And an estimator for the total number of elements in the population is

$$\boxed{\hat{N} = \sum_{h=1}^{L} \sum_{i=1}^{m_h} \sum_{j=1}^{n_{hi}} w_{hij}}$$

Now, if x_{hij} is the sample equivalent of X_{hij}, then an estimator for the population total X is,

$$\boxed{\hat{X} = \sum_{h=1}^{L} \sum_{i=1}^{m_h} \sum_{j=1}^{n_{hi}} w_{hij} x_{hij}}.$$

It now follows that in the context of a complex sampling design, the estimate of $D^s(z)$, which was previously

$$D^s(z) = \frac{1}{\hat{N}} \sum_{i=1}^{n} w_i x_i ,$$

becomes

$$\boxed{D^s(z) = \frac{1}{\hat{N}} \sum_{h=1}^{L} \sum_{i=1}^{m_h} \sum_{j=1}^{n_{hi}} w_{hij} x_{hij} = \frac{\hat{X}}{\hat{N}}}.$$

It should be clear that the two are equivalent, and thus the estimates of $D^s(z)$ do not need to be corrected for sampling. What does need to be corrected is the estimator of the variance of $D^s(z)$.

The *Stata 6.0 Reference Manual Su-Z*, pp. 69-70, shows that the an estimator for the variance of

$$\hat{R} = \frac{\hat{X}}{\hat{N}}$$

is

$$\boxed{\hat{V} = \sum_{h=1}^{L} \frac{m_h}{m_h - 1} \sum_{i=1}^{m_h} (d_{hi} - \bar{d}_h)^2}$$

where

$$d_{hi} = \sum_{j=1}^{n_{hi}} \frac{x_{hij} - \hat{R}}{\hat{N}} \qquad \text{and} \qquad \bar{d}_h = \frac{1}{m_h} \sum_{i=1}^{m_h} d_{hi} \, .$$

The four boxed items provide the basis for the dominance tests and the test statistics illustrated in part (1).