PREDICTION THEORY IN FINITE POPULATION SAMPLING: THE HORVITZ-THOMPSON ESTIMATOR AND ESTIMATES OF ITS VARIANCE

William G. Cumberland, U.C.L.A.
Richard M. Royall, Johns Hopkins University

1. INTRODUCTION

Prediction models in finite population sampling theory can reveal relationships which are essential for making inferences but which are often concealed in probability sampling analyses. Here we examine the Horvitz-Thompson estimator under some linear prediction models and compare the results with those derived through probability sampling theory under a probability-proportional-to-size (pps) sampling plan. Various estimators of variance are compared theoretically, using prediction models, and empirically, using a pps sampling plan to draw samples of \( n=32 \) repeatedly from six real populations.

2. THEORY USING THE PROBABILITY SAMPLING DISTRIBUTION

With each of the \( N \) population units two numbers are associated: the variable of interest \( y \) and an auxiliary variable \( x \) whose value is known. A sample of \( n \) units is chosen, leaving a set \( R \) of \( N-n \) non-sample units. The sample means are denoted by \( \bar{Y}_s \) and \( \bar{X}_s \), the non-sample means by \( \bar{Y} \) and \( \bar{X} \), and the population means by \( \mu_y \) and \( \mu_x \). The sampling fraction \( n/N \) is denoted by \( f \).

A probability sampling plan \( p \) specifies for every possible sample a the probability \( p(s) \) that \( s \) will be selected. The probability that unit \( i \) will be included is denoted by \( \pi_i = p(i \in s) \). When \( x_i > 0 \) is a measure of the size of unit \( i \), any plan \( p \) for which \( \pi_i = x_i/\sum_i x_i \) is proportional to \( x_i \) is called a probability-proportional-to-size (pps) sampling plan. We consider only sampling without replacement with a fixed sample size \( n \). In this case \( \pi_i = x_i/\sum_i x_i \) for all \( i = 1, \ldots, N \) (possible only when \( \sum_i x_i = \sum_i 1 \)). The Horvitz-Thompson estimator of the population total \( \mu_y = \sum_i y_i \pi_i \) is unbiased if \( \pi_i \) is known.

3. PERFORMANCE OF \( \hat{\mu}_T \) AND \( \hat{\sigma}_T^2 \)

The preceding theory treated the \( y_i \) as fixed constants. All probabilistic calculations referred to the distribution created by the probability sampling plan \( p \). For instance, the bias in \( \hat{\mu}_T \) was defined as \( E_p(\hat{\mu}_T - \mu) = \sum_s [N \bar{Y}_s - \mu] p(s) \), and the variance \( \sigma_T^2 \) was defined as \( \sum_s [N \bar{Y}_s - \mu] p(s) \). For instance, the bias in \( \hat{\mu}_T \) was defined as \( E_p(\hat{\mu}_T - \mu) = \sum_s [N \bar{Y}_s - \mu] p(s) \).

Theoretical comparison of these two variance estimators has proved difficult. Although all pps sampling plans for a given population and sample size \( n \) have the same inclusion probabilities, the joint inclusion probabilities \( \pi_{ij} \) can be different for different plans. This means that \( \sum_{s} \pi_{ij} \bar{Y}_s \) and \( \sum_{s} \pi_{ij} \bar{X}_s \) can be different for different plans. In the special case of perfect proportionality \( \pi_{ij} = \pi_i \) for every sample, so \( \sum_{s} \pi_{ij} \bar{Y}_s = \pi_i \bar{Y} \). In this case \( \hat{\sigma}_T^2 = 0 \), but \( \hat{\sigma}_T \) need not equal zero. This implies that \( \hat{\sigma}_T \) has greater sampling variance than \( \hat{\sigma}_C \) in populations where \( x_i \) is very nearly proportional to \( x_i \). It also shows that \( \hat{\sigma}_T \) can be negative in such populations.

It is less obvious but also well-known that \( \hat{\sigma}_C \) can take on negative values. On the other hand, \( \hat{\sigma}_T \) but not \( \hat{\sigma}_H \) has been shown to be non-negative for any sample chosen using some particular pps sampling plans (see Cochran 1977). The available empirical evidence favors \( \hat{\sigma}_C \) over \( \hat{\sigma}_H \), but this is based primarily on samples of \( n=2 \) from small populations (Rao and Singh 1973).

A third variance estimator sometimes recommended for its simplicity is one derived under the assumption of sampling with replacement. If we define \( d_i = y_i - \bar{x}_i \), then this estimator is \( \sigma^2 = \sum_i (d_i - \bar{y}_i)^2 / (n-1) \). Under a pps sampling plan \( \sigma^2 \) is biased. The bias is positive if the pps plan is more efficient than sampling independently with replacement and with probability \( x_i/\sum_i x_i \) of choosing unit \( i \) on each draw (Durbin 1953).

The pps sampling plan we have used in this study is the well-known one investigated by Hartley and Rao (1962). The sampling procedure consists of a random permutation of the units followed by a random-start systematic sample with step size \( \sum_i x_i \) through the interval \( (0, \sum_i x_i) \). Unit \( i \) is selected if \( \sum_{j<i} x_j < \sum_i x_i - \sum_{j<i} x_j \) for some of the systematic sample points \( u \). For this plan \( \sigma^2 \) has a positive sampling bias.

3. PERFORMANCE USING PREDICTION MODELS

The preceding theory treated the \( y_i \) as fixed constants. All probabilistic calculations referred to the distribution created by the probability sampling plan \( p \). For instance, the bias in \( \hat{\mu}_T \) was defined as \( E_p(\hat{\mu}_T - \mu) = \sum_s [N \bar{Y}_s - \mu] p(s) \) where \( \sum_s \) denotes the summation over all \( s \). The only random quantity was \( s \), the set of units chosen as the sample.

In many applications it is useful to regard the \( y_i \) as realized values of random variables \( (Y_1, Y_2, \ldots, Y_N) \). Knowledge about relationships among the \( y_i \)'s is already known to be a probability model for their joint distribution. In such cases a different approach to inference becomes available, one based on the probability model instead of the sampling plan. For instance, the bias can be defined as \( E_p(\hat{\mu}_T - \mu) = \sum_s [N \bar{Y}_s - \mu] p(s) \) where \( \sum_s \) denotes the summation over all \( s \). The only random quantity was \( s \), the set of units chosen as the sample.

Many applications it is useful to regard the \( y_i \) as realized values of random variables \( (Y_1, Y_2, \ldots, Y_N) \). Knowledge about relationships among the \( y_i \)'s is already known to be a probability model for their joint distribution. In such cases a different approach to inference becomes available, one based on the probability model instead of the sampling plan. For instance, the bias can be defined as \( E_p(\hat{\mu}_T - \mu) = \sum_s [N \bar{Y}_s - \mu] p(s) \) where \( \sum_s \) denotes the summation over all \( s \). The only random quantity was \( s \), the set of units chosen as the sample.
\[ \hat{\theta} = \frac{\sum (d_i/x_i)^2}{n(n-1)} \]

and substituting this estimate in (3.2) yields the least squares variance estimator \( \hat{\sigma}^2 \). This estimator is unbiased under model (3.1) and is closely related to \( \hat{\sigma}^2 \); \( \hat{\sigma}^2 \sim \hat{\sigma}^2 [1-\beta(2\bar{x}_s-c)/\bar{x}] \).

Royall and Cumberland (1978) have studied procedures for generating bias-robust variance estimators for linear statistics. These techniques produce estimators of \( \text{var} \hat{\theta}(\hat{\theta}_{HT}) \) which are unbiased under model (3.1) and which remain approximately unbiased when \( \text{var}(Y_i) \) is not proportional to \( x_i^2 \). One of these was studied in this study:

\[ \hat{\sigma}^2 = \frac{\sum (d_i/x_i)^2}{n(n-1)} \frac{\sum 2N\hat{\sigma}^2 / (1+g)}{2N\hat{\sigma}^2 / (1+g) / (n-1)} \]

The statistics \( \hat{\theta}_{HT}, \hat{\theta}_{VC}, \) and \( \hat{\theta}_L \) can also be viewed as estimators of (3.2) under model (3.1).

For the sampling plan we have chosen Hartley and Rao (1962) showed that for \( n \) \[ n \frac{1}{\hat{\sigma}^2} \text{var} \hat{\sigma}^2 = (n-1) \frac{(1+g)}{(1+g) / (1-N)} \]

(3.3) where \( \hat{\sigma}^2 \) is the least squares estimate of \( \sigma^2 \) and \( \hat{\sigma}^2 \) is the least squares estimate of \( \hat{\sigma}^2 \). Approximating the right hand side of (3.3) by (1-N1) and substituting in the formula for \( \text{var} \hat{\sigma}^2 \) leads to the following approximation:

\[ \text{var}(\hat{\theta}_{HT}) \approx \frac{N\hat{\sigma}^2}{(1+g)} / (n-1) \]

A similar substitution leads to an approximation for \( \text{var} \hat{\theta}_L, \text{var} \hat{\theta}_G \): \( \text{var}(\hat{\theta}_L) \approx \frac{N\hat{\sigma}^2}{(1+g)} / (n-1) \)

All the variance estimators \( \hat{\theta}_C, \hat{\theta}_D, \hat{\theta}_L, \) and \( \hat{\theta}_G \) have the same leading term, \( (N\hat{\sigma}^2) / (1+g) / (n-1) \), and are thus asymptotically equivalent (as \( n \to \infty \)). But when \( f \) is moderate there are important differences among the estimators. Under the model (3.1) \( \hat{\theta}_C, \hat{\theta}_D, \hat{\theta}_L, \hat{\theta}_G, \) are unbiased estimators of (3.2), while \( \hat{\theta}_{HT} \) is a biased estimator of (3.2). Defining the relative bias of an estimator by

\[ \text{rel bias}(\hat{\theta}) = \frac{\text{var}(\hat{\theta}_{HT}) / \text{var}(\hat{\theta}_L)}{\text{var}(\hat{\theta}_L)} \]

we have the following approximations under model (3.1):

\[ \text{rel bias}(\hat{\theta}_{HT}) \approx \frac{N\hat{\sigma}^2}{(1+g)} / (n-1) \]

where \( g = \beta^2 \). These approximations show that the relative bias in \( \hat{\theta}_C \) is small, for moderate \( f \), except that the four largest units have been removed from the extreme \( x \)-s and the sample units are chosen from the extreme \( x \)-s. The only variance estimator whose expected value depends on \( B \) is \( \text{var}(\hat{\theta}_{HT}) \), and this dependence in (3.2) yields the least squares variance estimator unbiased under the model (3.1) and is asymptotically unbiased under the model (3.6) in the sense that the relative bias is zero as \( n \to \infty \) and \( f \to 0 \). Since the other variance estimators have the same dominant term as \( \hat{\theta}_L \), all these estimators have the bias-robust property.

The flaw in \( \text{var}(\hat{\theta}_{HT}) \) is also robust, persisting under the general model (3.6). This is because the term \( \text{var}(\hat{\theta}_{HT}) \) depends on \( B^2 \beta^2 \hat{\sigma}^2 (\hat{\theta}_L) \), which causes the problem under model (3.1), remains unchanged under the more general model (3.6).

Failure of the condition \( E(\hat{\theta}_L) = \alpha x_i \) in (3.1) can seriously bias the estimator \( \hat{\theta}_{HT} \). For example, under the general 5th order polynomial regression model:

\[ EY_i = \sum \alpha j x_i^j, \text{var}(Y_i) = 0 \quad \text{if} j = 3 \]

the bias is

\[ \text{bias}(\hat{\theta}_{HT}) = \sum \alpha j x_i^j \]

where \( \alpha_j = \sum x_i^j / \sum x_i^j \). The bias vanishes when \( \alpha_j = 0 \) for \( j = 1,2,3 \), and we describe a sample satisfying these conditions as "balanced".

Failure of the condition \( E(Y_i) = \alpha x_i \) in (3.1) can seriously bias the estimator \( \hat{\theta}_{HT} \). For example, under the general 5th order polynomial regression model:

\[ EY_i = \sum \alpha j x_i^j, \text{var}(Y_i) = 0 \quad \text{if} j = 3 \]

the bias is

\[ \text{bias}(\hat{\theta}_{HT}) = \sum \alpha j x_i^j \]

where \( \alpha_j = \sum x_i^j / \sum x_i^j \). The bias vanishes when \( \alpha_j = 0 \) for \( j = 1,2,3 \), and we describe a sample satisfying these conditions as "balanced".

Failure of the condition \( E(Y_i) = \alpha x_i \) in (3.1) can seriously bias the estimator \( \hat{\theta}_{HT} \). For example, under the general 5th order polynomial regression model:

\[ EY_i = \sum \alpha j x_i^j, \text{var}(Y_i) = 0 \quad \text{if} j = 3 \]

the bias is

\[ \text{bias}(\hat{\theta}_{HT}) = \sum \alpha j x_i^j \]

where \( \alpha_j = \sum x_i^j / \sum x_i^j \). The bias vanishes when \( \alpha_j = 0 \) for \( j = 1,2,3 \), and we describe a sample satisfying these conditions as "balanced".

Failure of the condition \( E(Y_i) = \alpha x_i \) in (3.1) can seriously bias the estimator \( \hat{\theta}_{HT} \). For example, under the general 5th order polynomial regression model:

\[ EY_i = \sum \alpha j x_i^j, \text{var}(Y_i) = 0 \quad \text{if} j = 3 \]

the bias is

\[ \text{bias}(\hat{\theta}_{HT}) = \sum \alpha j x_i^j \]

where \( \alpha_j = \sum x_i^j / \sum x_i^j \). The bias vanishes when \( \alpha_j = 0 \) for \( j = 1,2,3 \), and we describe a sample satisfying these conditions as "balanced".

Failure of the condition \( E(Y_i) = \alpha x_i \) in (3.1) can seriously bias the estimator \( \hat{\theta}_{HT} \). For example, under the general 5th order polynomial regression model:

\[ EY_i = \sum \alpha j x_i^j, \text{var}(Y_i) = 0 \quad \text{if} j = 3 \]

the bias is

\[ \text{bias}(\hat{\theta}_{HT}) = \sum \alpha j x_i^j \]

where \( \alpha_j = \sum x_i^j / \sum x_i^j \). The bias vanishes when \( \alpha_j = 0 \) for \( j = 1,2,3 \), and we describe a sample satisfying these conditions as "balanced".

Failure of the condition \( E(Y_i) = \alpha x_i \) in (3.1) can seriously bias the estimator \( \hat{\theta}_{HT} \). For example, under the general 5th order polynomial regression model:

\[ EY_i = \sum \alpha j x_i^j, \text{var}(Y_i) = 0 \quad \text{if} j = 3 \]

the bias is

\[ \text{bias}(\hat{\theta}_{HT}) = \sum \alpha j x_i^j \]

where \( \alpha_j = \sum x_i^j / \sum x_i^j \). The bias vanishes when \( \alpha_j = 0 \) for \( j = 1,2,3 \), and we describe a sample satisfying these conditions as "balanced".

Failure of the condition \( E(Y_i) = \alpha x_i \) in (3.1) can seriously bias the estimator \( \hat{\theta}_{HT} \). For example, under the general 5th order polynomial regression model:

\[ EY_i = \sum \alpha j x_i^j, \text{var}(Y_i) = 0 \quad \text{if} j = 3 \]

the bias is

\[ \text{bias}(\hat{\theta}_{HT}) = \sum \alpha j x_i^j \]

where \( \alpha_j = \sum x_i^j / \sum x_i^j \). The bias vanishes when \( \alpha_j = 0 \) for \( j = 1,2,3 \), and we describe a sample satisfying these conditions as "balanced".
overestimate the variance. The averages for the other estimators are consistent with their being unbiased under \( \pi \text{ps} \) sampling.

The prediction theory outlined in Section 3 suggests that the performance of the variance estimates will depend strongly on \( \bar{x}_s \). The bias incurred by \( \text{VHT} \) when \( \text{EM}(\text{VHT} - \text{VHT}_0) \) contains a quadratic term also depends on \( \bar{x}_s \). To examine performance as a function of \( \bar{x}_s \), we arranged the 1000 samples from each population in order of increasing value of \( \bar{x}_s \). We then grouped the samples in 20 sets of 50, so that the first group contains the 50 samples with the smallest \( \bar{x}_s \), the next group the samples with the next 50 smallest \( \bar{x}_s \), etc. For each group we calculated the average values of \( \bar{x}_s \), the average error, the mean square error (mse), and the averages of each of the five variance estimates, \( \text{VC}_0 \), \( \text{VC}_L \), \( \text{VHT} \) and \( \text{VYG} \). We then plotted the average errors and the values of \( \text{mse}^{1/2}, \text{VC}_0^{1/2}, \text{VC}_L^{1/2} \) and \( \text{VHT}^{1/2} \) against the average values of \( \bar{x}_s \). The trajectory showing the average error in (3.8) resulting from quadratic curvas \( \bar{x}_s \) is labelled error. Those showing \( \text{VC}_0^{1/2}, \text{VC}_L^{1/2} \), etc., are labelled C, D, etc. Figures 1-6 show the results.

The performance of VHT was so dramatically different from the others that we have made separate figures, 7-12, for comparing VHT with VYG.

Probability sampling theory assures us that \( \text{VHT} \) is unbiased in \( \pi \text{ps} \) sampling. Table 2 confirms this. However, as in Figures 1-6 suggests that the larger the bias defined with respect to the prediction model and conditioned on \( s \) is more appropriate for inference from a given sample. In four populations (Cities, Counties 60, Counties 70 and Hospitals) there is no clear bias in \( \text{VHT} \) as a function of \( \bar{x}_s \). The positive and negative biases at the extremes cancel each other when the results are averaged over all samples. Thus averaging over the sampling distribution conceals an important property of the estimator, its bias in samples not achieving good \( \pi \)-balance. The bias seen in \( \text{VHT} \) as a function of \( \bar{x}_s \) can be explained in terms of curvature of the true regression function. The term in the bias of \( \text{VHT} \) found in (3.8) resulting from quadratic curvas \( \bar{x}_s \) is \( \text{N}(\text{mse}, \text{mse}) \). Hence for populations like Counties 70 with a convex regression (slope increasing with \( x \)), the bias should be an increasing function of \( \bar{x}_s \). For populations like Counties 60 and Hospitals where the regression is concave, the bias should be decreasing linear function of \( \bar{x}_s \). This bias will be near zero when \( \bar{x}_s = c \). This is precisely the behavior suggested by the error curves in the figures.

Note that \( \pi \text{ps} \) sampling has not provided adequate \( \pi \)-balance to protect \( \text{VHT} \) against bias in these populations. This is particularly true of Counties 60, where the 10-20% of samples in which \( |\bar{x}_s - c| \) was greatest produced biases which are large relative to the (mse). It illustrates an important point: \( \text{VHT} \) sampling produces \( \pi \)-balance "on the average" does not imply that we can make inferences as if each sample were approximately \( \pi \)-balanced.

The variance estimates behave as predicted under model (3.1). The curves for \( \text{VC}_0 \), \( \text{VC}_L \) and \( \text{VYG} \) all tend to decrease as \( \bar{x}_s \) increases, and all three estimators track the mse curve reasonably well except in cases where large bias influences the mse. The difference between \( \text{VC}_0 \) and the others increases as \( \bar{x}_s \) increases in all the populations.

To examine the performance of \( \text{VHT} \) we plotted the average values of \( \text{VHT} \) and \( \text{VHT}_0 \) against the average values of \( \bar{x}_s \) in Figures 7-12. Although \( \text{VHT} \) and \( \text{VHT}_0 \) are unbiased in expectation with respect to a \( \pi \text{ps} \) sampling plan, prediction theory reveals that \( \text{VHT} \) can have a serious bias when \( \bar{x}_s \), and under the model (3.1) this bias is a linear increasing function of \( \bar{x}_s \). When \( \bar{x}_s < x^* \), given in (3.5), the bias will be so large that the expected value under (3.1) of \( \text{VHT} \) is negative. This will also be true of \( \text{VHT}_0 \), if \( \text{VHT}_0 \) is an adequate approximation. The theory outlined in Section 3 suggests that for larger \( \beta^2/\sigma^2 \), the more severe the bias in \( \text{VHT} \). For each population the least-squares estimates of \( \beta \) and \( \sigma^2 \) under (3.1) based on the entire population were used to estimate \( \beta^2/\sigma^2 \) (see Table 1). These values were used to estimate \( x^* \) (3.5) where \( \text{EM}(\text{VHT}) = 0 \). The points \( x^* \) are on the x-axes in Figures 7-12. The extent to which prediction theory has correctly summarized the performance of \( \text{VHT} \) in these populations is remarkable. In every population the \( \text{VHT} \) line increases linearly with \( \bar{x}_s \). In the populations in which the \( \text{VHT} \) line becomes negative (Cities, Counties 1960, Counties 1970, and Sales), it crosses the axis near \( x^* \). In those populations with the largest values of \( \beta^2/\sigma^2 \) (Countries 1960 and Sales) performance of \( \text{VHT} \) is worst. Cancer and Hospitals, having the smallest values of \( \beta^2/\sigma^2 \), are the only two populations for which the \( \text{VHT} \) line is non-negative. In both of these populations, \( x^* \) is smaller than the smallest average \( \bar{x}_s \) value plotted. As theory predicts, the lines for \( \text{VHT} \) and \( \text{VHT}_0 \) cross when \( \bar{x}_s = x^* \) in every population.

5. CONCLUSIONS

The variance estimator \( \text{VHT} \) is generally recommended over \( \text{VHT}_0 \). The present study provides strong theoretical and empirical support for this advice. It also suggests that \( \pi \text{ps} \) and \( \pi \text{ps} \) deserve consideration as practical alternatives. These have two important advantages over \( \pi \text{ps} \)—they are always non-negative and they do not require that the joint inclusion probabilities \( \pi_{ij} \) be available.

We believe these results show again that finite population inferences should be based on prediction models, not on the probability sampling distribution. They help to clarify how probability sampling can contribute to robust inference with \( \text{VHT} \) by producing samples which are approximately \( \pi \)-balanced. But they illustrate the danger in making inferences which are not conditioned on the sample is actually observed.

REFERENCES

Table 1: Study Populations

<table>
<thead>
<tr>
<th>Population</th>
<th>x</th>
<th>y</th>
<th>T</th>
<th>$b^2/s^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cities</td>
<td>125 U.S. cities</td>
<td>population 1960</td>
<td>population 1970</td>
<td>35.691x10^6</td>
</tr>
<tr>
<td>Counties 60</td>
<td>304 counties in N.C., S.C. and Ga.</td>
<td>households 1960</td>
<td>population 1960</td>
<td>10.007x10^6</td>
</tr>
<tr>
<td>Hospitals</td>
<td>393 hospitals from national sample</td>
<td>number of beds</td>
<td>patients discharged</td>
<td>320159</td>
</tr>
<tr>
<td>Sales</td>
<td>327 U.S. Corporations</td>
<td>gross sales 1974</td>
<td>gross sales 1975</td>
<td>667.88x10^9</td>
</tr>
</tbody>
</table>

$^*$b and $s^2$ are weighted least-squares estimates of $\beta$ and $\sigma^2$ under the model (3.1).

Table 2: Results for 1000 $\psi_n$ samples of $n = 32$ (Horvitz-Thompson estimator)

<table>
<thead>
<tr>
<th>Population</th>
<th>Average Error</th>
<th>$(\hat{T}_{HT} - T)^2$</th>
<th>$V_C$</th>
<th>$V_D$</th>
<th>$V_L$</th>
<th>$V_{YG}$</th>
<th>$V_{HT}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cancer</td>
<td>4.2</td>
<td>502</td>
<td>539</td>
<td>479</td>
<td>464</td>
<td>489</td>
<td>487</td>
</tr>
<tr>
<td>Cities (millions)</td>
<td>.02</td>
<td>1.26</td>
<td>1.46</td>
<td>1.18</td>
<td>1.13</td>
<td>1.23</td>
<td>1.24</td>
</tr>
<tr>
<td>Counties 60 (thousands)</td>
<td>.155</td>
<td>110</td>
<td>124</td>
<td>108</td>
<td>105</td>
<td>109</td>
<td>107</td>
</tr>
<tr>
<td>Counties 70 (thousands)</td>
<td>15.4</td>
<td>267</td>
<td>318</td>
<td>271</td>
<td>269</td>
<td>266</td>
<td>269</td>
</tr>
<tr>
<td>Hospitals (thousands)</td>
<td>-.42</td>
<td>14.0</td>
<td>14.9</td>
<td>14.1</td>
<td>13.9</td>
<td>14.0</td>
<td>14.0</td>
</tr>
<tr>
<td>Sales (billions)</td>
<td>.10</td>
<td>13.0</td>
<td>14.1</td>
<td>12.7</td>
<td>12.3</td>
<td>12.9</td>
<td>13.1</td>
</tr>
</tbody>
</table>

(Average in 1000 samples)$^\dagger$