APPENDIX K

SAMPLE ADEQUACY ASSESSMENT SOFTWARE

INTRODUCTION

The sample adequacy assessment provides a comparison of the actual number of samples to the required number of samples as defined in Chapter VII and Appendices C and D of this Manual. Analysis of the reports provided will indicate where more samples are required or where some samples may possibly be deleted (see explanations for columns 21 and 25 of the reports). The number of existing and required samples is determined for each functional system for each AADT volume group by rural, small urban and individual urbanized area or collective urbanized area for the standard sample panel and for each nonattainment area donut area sample panel.

To obtain the Sample Adequacy reports, select *Analysis/Sample Management/Adequacy* from the menu bar in the HPMS Submittal Software. Either the reports can be sent directly to the printer by pressing the *Print* button or they can be viewed on the screen by pressing the *Preview* button.

Any data file to be processed for sample adequacy assessment should contain CLEAN, EDITED, and CORRECTED data and have volume groups and expansion factors assigned.

The standard sample adequacy assessment output consists of several reports based on Appendix D (Sample Size Estimation Procedures).

- Sheet 1 of Report ADEQ-01 contains a summary of the current standard sample and universe data by functional system and AADT volume group.

- Sheet 2 of Report ADEQ-01 contains the Appendix D calculations for the estimated number of required samples, the difference from the current number of unique standard samples as well as the estimated (or actual) universe records and length, for the same functional system and appropriate volume group strata.

There are Sheet 1/Sheet 2 report pairs as indicated above for each of the following:

- Rural

- Small Urban

- Each Urbanized Area and/or Collective Grouping of Urbanized Areas

The donut area sample adequacy assessment reports follow the standard sample adequacy reports, if any donut areas exist. This report consists of two sheets for each donut area:

- Sheet 1 of Report ADEQ-02 contains rural and small urban universe counts and length values summarized by the appropriate functional systems and volume groups.

- Sheet 2 of Report ADEQ-02 contains rural and small urban donut area sample information summarized by volume group for the combined rural and small urban minor arterial and combined rural major collector and small urban collector systems.

REPORT FIELD DESCRIPTIONS

This section describes the contents of each column in the reports and should be read and understood before the reports are used.

NOTES:

Some of the subtotals and totals may not equal the sum of the entries making up that total. This is because the line-by-line values are rounded as they are printed, while the actual values are accumulated for the totals and are not rounded until the totals are printed. The differences should be small.

The title lines identify which Appendix C table was used for the confidence levels and precision rates (see Columns 26 and 27 of Sheet 2 of any report). The title line messages and the conditions under which each will appear are as follows:

Title Line Message	Condition Under Which Each Will Appear
Using Appendix C, Table C-1	For rural areas.
Using Appendix C, Table C-2	For small urban areas.
Using Appendix C, Table C-3/ Individual Areas (<3)	For urbanized areas with population less than 200,000 (rural/urban code is 3), where there are fewer than 3 individual urbanized areas in the State.
Using Appendix C, Table C-3/ Individual Areas (3 or More)	For urbanized areas with population less than 200,000 (rural/urban code is 3), where there are 3 or more individual urbanized areas in the State.
Using Appendix C, Table C-3/ Collective Areas	For collective urbanized areas. These areas will be reported in column 4 with an area code of 901 to 909.
Using Appendix C, Table C-4	For urbanized areas with population greater than 200,000 (rural/urban code is 4), or for urbanized areas with a population less than or equal to 200,000 (rural/urban code is 3) which are also in a nonattainment area.
Using Appendix C, Table C-5	For donut areas of nonattainment areas.

EXPLANATION OF COLUMNS IN SHEETS 1 AND 2 OF REPORTS ADEQ-01 AND ADEQ-02

HPMS DATA

These columns classify the data from both the sample and the universe. Sample column comments apply to both standard and donut area samples. If there is a difference, it is explained.

Column 1 – Functional System: This column contains HPMS Item 17, functional system code. All local (Item 17 = 9 or 19) and rural minor collector (Item 17 = 8) functional system records are excluded from this analysis.

For a donut area, this column will contain rural minor arterial (Item 17 = 6) followed by small urban minor arterial (Item 17 = 16), then by a summary Item 17 = 6, 16 line; this is followed by rural major collector (Item 17 = 7), then small urban collector (Item 17 = 17) and a summary Item 17 = 7, 17 line.

Column 2 – R/U Code: This column contains HPMS Item 13, rural/urban designation code.
This code is used as the basis for the precision levels used in the Appendix D calculations.

Column 3 – Volume Group: This is either the standard sample AADT volume group identifier, Item 32 (codes 1-13 in Report ADEQ-01) or the donut area sample AADT volume group identifier, Item 31 (codes 1-5 in Report ADEQ-02). The text that follows describes other entries that may be coded in addition to the numeric volume group codes:

- "**G**" as a Suffix after the Volume Group: Indicates that the values on this line are for Grouped length data (Item 9 = 1). The AADT volume group is known (it is a required data item) but the actual AADT is not known since it is likely assigned (or coded zero) for the group of sections. On Sheet 1 of the report, these values will be reported separately; while on Sheet 2 of the report, an estimate of the number of sections in the group will be included with the actual number of sections (see Column 16). Column 3 also appears on Sheet 2 of Reports ADEQ-01 and ADEQ-02 for control purposes. The "**G**" suffix print line appears on Sheet 2 (even though no data appears on these lines) to enable easy comparison between Sheets 1 and 2.

- **TOTAL** – For standard samples, this indicates totals for a functional system.

  For donut area samples, **TOTAL** indicates totals for a combined functional system (Item 17 = 6,16 or Item 17 = 7,17).

- **GRTOT** – For standard samples, this indicates the grand total for the sheet (rural, small urban, individual urbanized area, or collective urbanized group).

  For donut area samples, **GRTOT** indicates the grand total for the donut area of a nonattainment area.

Column 4 – Urbanized Area Code: This will be zero (dash) for rural and small urban entries, and will identify the individually sampled urbanized areas, Item 15 (as taken from Appendix B). Collective urbanized areas will be identified as "901" for group 1 (Item 14 plus 900); "902" for group 2; etc.

Report ADEQ-02 contains the nonattainment area code for the donut area.

SAMPLE DATA

These columns include summaries only from those records that are samples.

Column 5 – Number of Samples: This is a count of the number of samples. The resulting records are then used to develop Column 10, the sample AADT coefficient of variation; this is also the "current" value used to calculate the number of required samples in Columns 18 and 22.

Column 6 – Unexpanded Length: This is the accumulated sample length of the total number of sections contained in Column 5.

Column 7 – Average Section Length: Obtained by dividing Column 6 by Column 5, and is for information purposes only.

Column 8 – Expanded Length: The sum of the expansion factor times the length for all of the sections in the volume group (Column 5). This value should agree with the total universe length (Column 11). If some universe records in a volume group have been grouped, they are summarized in the line below the ungrouped sections. The two Column 11 lengths must be summed to make a valid check with Column 8 for the particular volume group.

Column 9 – Expansion Factor: This column contains the standard expansion factor (Item 49) or the donut area expansion factor (Item 48). If all records of a volume group do not have the same expansion factor, the first expansion factor encountered will be contained in this column.

Column 10 – AADT Coefficient of Variation (C.V.): This is computed by dividing the AADT standard deviation by the mean AADT for each volume group using only the samples identified as unique (Column 5). The calculated C.V. is the "C" value used in the Appendix D formula for calculating the required number of samples in Column 18.

If there are no samples (or only one) in the volume group with AADT coded, the coefficient of variation cannot be calculated. Under this circumstance, predetermined coefficients are used (if the predetermined (default) value is used, a "T" will follow the value). This mainly occurs where the universe AADT data has a volume group that is not sampled or all samples in the volume group contain the same AADT, but could also happen when there is undersampling. The same rule applies to the universe coefficient in Column 14. When the computed C.V. value is less than 0.005, the default values will also be used. This usually occurs when the difference in AADT values among records is very small.

If a sample record has no AADT coded, it will be excluded from the C.V. computation and an asterisk (*) will follow the C.V. value.

If both a "T" (default table used) and an "*" (one or more AADT values of zero) apply to a particular volume group, only the "T" notation will appear.

UNIVERSE DATA

These columns contain a summary of all HPMS records, including samples.

Column 11 – Length: The total length for the volume group. Lengths of grouped records (Item 9 = 1) are contained on the lines where Column 3 has the "**G**" suffix.

Column 12 – Number of Sections: A count of the number of HPMS records representing the volume group. Grouped length records (Item 9 = 1) are contained on the lines where Column 3 has the "**G**" suffix.

Column 13 – Average Section Length: Column 11 divided by Column 12. The ungrouped average value is used to calculate the estimated number of sections for the grouped sections, if needed–see the explanations for Columns 15 and 17.

Column 14 – AADT Coefficient of Variation (C.V.): As described for Column 10, but in this case all universe records that contain AADT are used.

If a universe record has no AADT coded, it will be excluded from the C.V. computation and an asterisk (*) will follow the C.V. value.

If there is only one universe record (or only one with an AADT value), the C.V. will be taken from default values contained in the software and a "T" will follow the C.V. value.

The calculated or default C.V. is the "C" value used in the Appendix D formula for calculating the number of samples required (Column 22). Grouped length records (Item 9 = 1) identified by a "**G**" suffix in Column 3 and total lines (**TOTAL** and **GRTOT**) will not have a coefficient of variation.

FROM UNIVERSE DATA

Estimations (where applicable) are made for the number of records in each volume group based on what is known from the existing data.

Column 15 – Estimated Number of Sections: This column contains the actual number of universe records in each volume group from Column 12 unless records for grouped length (Item 9 = 1; "**G**" suffix in Column 3) were included. Where there were grouped length records, the estimated number of sections is computed as the sum of the actual number of sections in Column 12 plus the total length (Column 11) for the grouped length records for that volume group divided by the average section length (Column 13) for those sections that were not grouped. Note that for this calculation, if the average section length exceeds 10.00 miles per section, the value 10.00 will be used.

This value becomes "N" in the Appendix D formula for calculating the required number of samples.

Column 16 – "EST" if Estimate Made: If Column 15 contains an estimate of the number of sections as described under Column 15 (from grouped length records), the abbreviation "EST" appears in this column for the volume groups involved. If "EST" is not present, the actual number of sections was brought forward from Column 12 (universe data) and placed in Column 15.

Column 17 – Combined Length: This column contains the combined universe length for each volume group from the two possible Column 11 entries; the one for the non-grouped length records plus the one for the grouped length records, if any existed.

The rest of the columns contain calculated values and values used in the calculations.

FROM THE SAMPLE AADT COEFFICIENT OF VARIATION (C.V.)

Column 18 – Number of Required Samples: This is the result of using the C.V. developed from the sample data (Column 10), the number of sections available for sampling (Column 15), and the confidence level value and precision rate contained in Columns 26 and 27 (from Appendix C) as input to the formula in Appendix D.

An asterisk (*) following the number of required samples in this column indicates that the required number of samples calculated from the Appendix D formula results in an estimated expansion factor greater than 100.000. The number of samples contained in Column 18 includes the original calculated value plus those contained in Column 19 so that the expansion factor would not exceed 100.000. For example, if the computed required number of samples resulted in an expansion factor of 110.000, the number of samples contained in this column has been increased by the number of samples indicated in Column 19 to keep the expansion factor at 100.000 or less. The analyst may want to consider increasing the number of samples even further above this value to allow for future changes.

This calculation is made in addition to the one based on the universe C.V. (Column 22) for comparison purposes and to provide a choice to the State based upon where the more accurate AADTs are encoded. If the State believes that the universe AADTs are accurate throughout, then the Column 14 C.V. should be used and Column 22 is considered a reasonable estimate of the sample requirement.

NOTE: Once a State elects to use either the sample estimate (column 18) or the universe estimate (column 22), the entire sample panel is to be based on that estimation criteria (sample or universe).

Column 19 – Number of Factor Samples: This column contains the number of samples that were added to the calculated number so that the estimated expansion factor would not exceed 100.000. If the number of required samples as originally calculated for Column 18 resulted in an expansion factor greater than 100.000, an asterisk (*) will appear in Column 18. Column 19 will indicate how many additional samples were required to keep the expansion factor from exceeding 100.000.

Column 20 – New Expansion Factor: This column contains the estimated expansion factor that would result from the number of samples shown in Column 18. If there is an asterisk (*) in Column 18, the expansion factor is the result after the number of samples shown in Column 19 has been added.

Column 21 – Difference: Required Samples Minus the Current Samples: Column 18 minus Column 5. A minus value indicates that the volume group has more than enough samples to be statistically sound. While differences in the range of + or - 10 percent may be generally ignored, several other considerations are needed before making any final decisions about, or changes to, the State's sample panels:

1. A prerequisite to doing anything is to ensure that the AADT data are up to date and accurate, and that all applicable records have been included in the analysis. Otherwise, this analysis cannot be considered a true assessment of the State's sample adequacy.

2. A comprehensive report of intended actions for <u>reductions</u> in the number of samples is to be submitted to FHWA Headquarters, HPPI-20 **BEFORE** any such action actually takes place. Random deletion of samples is a must in any such plan. The reduction plan will be evaluated at FHWA Headquarters and appropriate remarks will be returned via the FHWA field offices.

   Note: Sample section additions may take place at any time without FHWA evaluation.

3. A volume group must contain a minimum of three samples or all that are available if there are less than three universe records in the volume group.

4. The expansion factor should not be greater than 100.000 to ensure a statistically sound sample panel, and probably should be kept at even lower levels to provide for change over time.

5. If a State is using the HPMS Analytical Process, HERS, or is using the HPMS sample panel for other purposes, it may want to consider using higher confidence levels and/or precision rates than were used in these calculations; this will result in a larger number of required samples, of course.

6. FHWA recommends that the number of samples not be reduced below a 10-percent difference, if more than that difference already exists (i.e., if 32 samples are required, 35 or so should be retained if a volume group already contains more than 35). This will provide for movement of samples and universe records into other volume groups over time.

7. The State should examine its AADT trends, the shifting of samples from one volume group to another over the years, the future expectations of AADT change, and other factors concerning AADT before making decisions concerning reductions in the number of samples.

8. This summary is an <u>ESTIMATE</u>. The State should do its own analysis, or at a minimum, ensure that this summary is a reasonable estimate of the State's sample panel requirements (particularly where estimates for the number of volume group records have been made in Column 15).

    Additional information about sample panel reduction is contained in Chapter VII.

9. Unsampled or undersampled volume groups must also be addressed, under any sample panel review.

FROM THE UNIVERSE AADT COEFFICIENT OF VARIATION (C.V.)

Column 22 – Number of Required Samples: This is the result of using the C.V. developed from the universe data (Column 14), the number of sections available for sampling (Column 15), and the confidence level value and precision rate contained in Columns 26 and 27 (from Appendix C) as input to the formula in Appendix D.

An asterisk (*) could appear in this column; see explanation under the description for Column 18.

This calculation is made in addition to the one based on the sample C.V. (Column 18) for comparison purposes and to provide a choice to the State based upon where the more accurate AADTs are encoded. See the discussions under Columns 18, 19, 20, & 21 for more details.

Column 23 – Number of Factor Samples: This column contains the number of samples that were added to the calculated number so that the estimated expansion factor would not exceed 100.000 (see Column 18 for explanation). If the number of required samples as originally calculated for Column 22 resulted in an expansion factor greater than 100.000, an asterisk (*) will appear in Column 22. This Column (Column 23) will indicate how many additional samples were required to keep the expansion factor from exceeding 100.000.

Column 24 – New Expansion Factor: This column contains the estimated expansion factor that would result from the number of samples shown in Column 22. If there is an asterisk (*) in Column 22, the expansion factor is the result after the number of samples shown in Column 23 has been added.

Column 25 – Difference: Required Samples Minus the Current Samples: Column 22 minus Column 5. See the discussions under Columns 18, 19, 20, & particularly 21 for more information on the uses, ramifications, and cautions concerning these differences.

Column 26 – Standard Value for the Confidence Level: The standard values come from statistical handbooks and are used as "Z" in the Appendix D formula for calculating the number of required samples:

Value	Confidence Level
1.645	90 Percent
1.282	80 Percent
1.04	70 Percent

The confidence levels come from the tables in Appendix C. The State may wish to raise these levels if it plans to use the HPMS sample panel for its own purposes and requires higher levels of confidence.

Column 27 – Precision Rate Desired: The precision rates come from the tables in Appendix C and are used as "d" in the Appendix D formula for calculating the number of required samples. The State may wish to raise these rates if it plans to use the HPMS sample panel for its own purposes and requires more precision.