

RECEIVED

SEP 24 4 57 PM '01

POSTAL RATE COMMISSION
OFFICE OF THE SECRETARY

Rural Carrier Cost System (RCCS)

Statistical and Computer Documentation

(Source Code and Data on CD-ROM)

USPS-LR-J-13/R2001-1

This document is a Category 1 library reference. It replaces USPS-LR-I-17/R2000-1. It relates to the City Carrier Cost data reporting system, which is the subject of the testimony of witness Harahush, USPS-T-5.

TABLE OF CONTENTS

Overview	2
Use of RCCS Data in Cost Attribution	2
Statistical Study Design	3
First Stage Sample	3
Second Stage Sample	3
Third Stage Sample	3
System Flowchart	4
Creating the Sample Frame	5
Selecting the First Stage Sample	7
Stratification	7
Sample Allocation	8
Sample Selection	8
Joint Scheduling	10
Partitioning the Sample	11
Administering Tests in the Field	12
Downloading the Sample	12
Administrative Functions	12
Collecting the Data	12
Transmitting the Data to San Mateo	13
Data Validation and Editing	14
Manual Editing	14
Computerized Editing	14
Estimation	15
Quarterly Estimation	19
Annual Volumes	22
File Layouts	23
Rural Route (Header) File	23
Rural Shape File	24
Rural Mailcode File	25
Rural Sample Selection File	28
Rural Mcode File	30
Rural Z File	31

OVERVIEW

The Rural Carrier Cost System (RCCS) is a continuous, ongoing statistical study, or probability sample of rural carrier route-days. For each selected route-day, a sample of mail is selected, and for each selected mailpiece, the class, subclass, and compensation category of mail is recorded directly into a portable microcomputer using the Computerized On-Site Data Entry Systems (CODES) software.

The RCCS gathers data for distributing major portions of carriers' salaries, benefits and related costs to the various categories of mail for postal rate-making purposes. Accrued carrier costs, available from payroll data, are total amounts and are not generally associated with any particular class of mail or service. Therefore, special methods are needed to determine the costs associated with the various products and services.

Rural carrier activity consists of delivering mail to and collecting mail from customers located on rural routes. In addition it includes certain activities such as providing special services, collecting postage, and selling stamps. A rural carrier conducts almost all of the activities of a post office.

Rural delivery is organized and operated in terms of individual routes. Rural routes are divided into two broad categories, depending on the way the carrier is paid. Most rural routes are evaluated routes, that is, the rural route is evaluated in terms of time standards, and the carrier is paid a salary based on the evaluated time. Evaluated routes include the H, J, and K route types. The evaluated time is developed from route factors such as route length, boxes served, and number of letters delivered. Other routes include mileage routes, type M, paid on the basis of mileage on the route, and auxiliary routes, type A, paid on the basis of hours worked.

Use of RCCS Data in Cost Attribution

Total accrued costs for rural carriers are summarized in Cost Segment 10 (CS 10). The costs are divided into separate components for evaluated routes and other routes, based on payroll records.

The route factors are measured during the National Rural Mail Count, which is usually conducted biannually. During the National Rural Mail Count, all mail for a large proportion of the rural routes is counted, and time measurements for other factors are measured. Thus factors related to volume (volume variable cost drivers) and factors independent of volume (fixed cost drivers) are measured during the National Rural Mail Count.

The volume variable costs of rural carrier workhours are determined by/from a variability analysis developed in accordance with the evaluated time and factors of workload derived from the rural routes participating in the National Mail Count. Volume variable costs are determined for each of the evaluated and other route components of Cost Segment 10.

Data from the Rural Carrier Cost system are used to distribute volume variable costs across mail subclasses and special services. The delivery portion of the Rural Carrier Cost system (data collected via the CODES data collection system) provides the subclass and special service data for mail delivery costs. The PS Form 2848 portion of the Rural Carrier Cost system provides the subclass and special services data for the distribution of volume variable mail collection costs.

STATISTICAL STUDY DESIGN

The universe under study in RCCS is all mail being delivered on rural routes. A stratified, three stage sample design is used for RCCS. The details for each of the stages are listed below.

First Stage Sample (Route)

The first stage sample is a stratified random sample of route-days. Every rural route is assigned to one of two strata based upon whether the route is in a district having 20 or fewer routes or if it is in a district with more than 20 routes. Within each stratum, routes are geographically ordered, and a systematic random sample of routes is selected. Possible delivery dates (every Monday through Saturday, excluding holidays) are randomized, and systematically assigned to selected routes, to determine the route-days, or first stage sample units to be enumerated. This selection process ensures both geographic and temporal dispersion of the sampled route-days, and helps control workload at the District level.

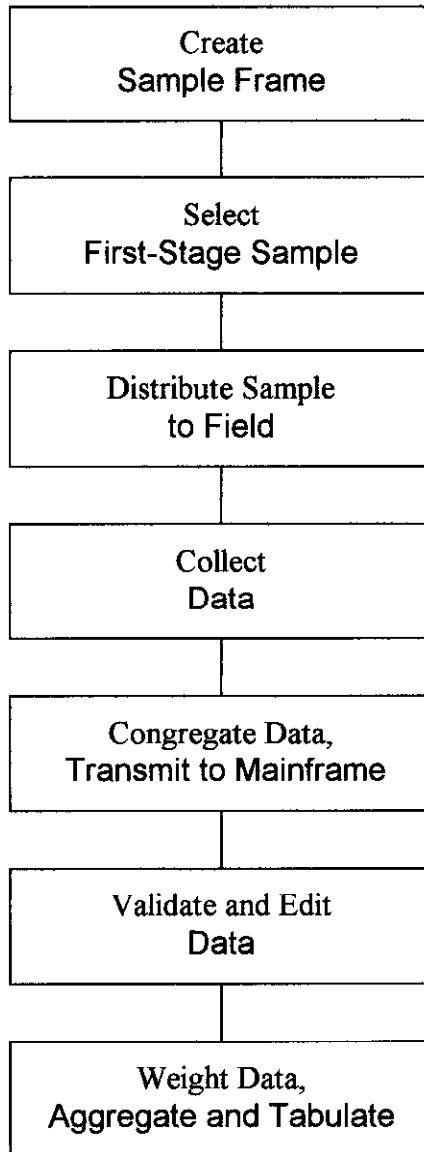
Second Stage Sample (Office)

The second stage sample unit is the office, when the selected route serves one or more intermediate offices. In that case, one office served by the route is randomly selected. However, the majority (about 96%) of the routes do not serve intermediate offices, and are completely enumerated at the second stage.

Third Stage Sample (Mailpiece)

The third stage sampling unit is a mailpiece. Parcels and accountables are completely enumerated. A systematic sample of letters and flats is selected. The data collector determines the skip interval ("S") to be used – typically 10 – and the CODES software generates a random number "R", between one and "S". The data collector selects the "Rth" piece, and every "Sth" piece thereafter. The recommended skip interval is 10. Data collectors are allowed to change skip intervals as the need arises. The skip interval used is stored on each mailpiece record.

SYSTEM FLOWCHART



CREATING THE SAMPLE FRAME

The sampling frame, or Rural Master Frame, is extracted from the Rural Payroll file, which contains a list of all rural carrier routes.

1. The Minneapolis Finance department runs an extraction program once each quarter to generate a current list of rural carrier routes. This extract list of rural carrier routes is the Rural Master Frame for the Postal Quarter. Before running the Sample Selection Process, Headquarters must verify that the extract file `AAQMNN.PS700T01.RURLEXTR.PQqFYyy` for the specific quarter exists.

After the creation of the frame, but before the beginning of sample selection a date file is produced. As an input, the first line of the previous year's date file is used. That line contains the sample percentages for the same quarter of the previous year. The output is last year's sample percentages with this year's delivery days.

The date file (a file that contains sampling percentages, the number of delivery days in the quarter, the day of week and the date for each delivery day in a quarter) is created by executing the `ALDRAN.FYyyyy.MASTER.SAMPLE.PROD.JOBS(DATEFIL1)`. The date file serves two functions 1) to allocate the sample evenly throughout the quarter and 2) to give the sample selection program the sampling percentages (the first record of the output date file). Before executing `DATEFIL1`, RPW personnel prepare and execute the RPW Date File job "`HSF.HQ000D01.CARDDATA(HSF020DP)`".

The input and output file(s) for executing `DATEFIL1` follow:

INPUTS

1. RPW Date File - `HSF.HQ000D01.CARDDATA(HSF020SA)`
2. Date File - `ALDRAN.FYyyyy.PARMLIB(DATE.PQq)`
3. SAS Program - `ALDRAN.CCSREDES.PDS.SRC(ALD080C1)`

OUTPUTS

1. Date File - `ALDRAN.FYyyyy.PARMLIB(DATEPQq)`

The sample sizes for the two strata are determined by the two Rural sampling percentages in the first record of the date file. Before the program is executed two parameters need to be changed:

Line 11400 PQ = (one digit 1-4)

Line 11500 FY = (two digit year)

Rural sampling percentages are reviewed in the program `ALDRAN.FYyyyy.MASTER.SAMPLE.PROD.JOBS(MUNIVSIZ)`. That program calculates the total routes in each of the two strata and obtains sample sizes for analyst review. The analyst utilizes output from `MUNIVSIZ` to determine sample size for the current quarter and changes the date file sample percentages in preparation for the running of the next program `SMPLSEL`.

The sample is selected by executing the program ALDRAN.FYyyyy.MASTER.SAMPLE.PROD.JOBS(SMPLSEL). SMPLSEL utilizes the revised date file as an input.

Below is an explanation of the inputs and outputs of SMPLSEL. In the next section, "Selecting the First Stage Sample," more detail of the program SMPLSEL is explained.

Before executing the program a random 9 digit seed must be chosen from a random number table. This seed is used to insure the randomness of the dates and routes to be selected.

INPUTS

1. Rural Payroll Extract File - AAQMNN.PS700T01.RURLEXTR.PQqyy
2. Date File -ALDRAN.PS000D01.DATE.PQq
3. Ziptable File – HSFRAN.PS001D01.ZIPTABLE.ORFEO
4. Totals File - &&CRTOTALS (temporary SAS data set with the correct sampling percentages for each stratum)

OUTPUTS

1. Sample File - ALDRAN.PS400T01.RURAL.PQqFYyy
2. Sample File Vault copy - ALDRAN.PS400V01.RURAL.PQqFYyy
3. Rural Master Frame File - ALDRAN.PS401T01.RURAL.PQq.FYyy
4. Rural Master Frame File Backup - ALDRAN.PS700T01.RURLEXTR.BKUP.PQqFYyy
5. Rural Master Frame File with Packed fields transformed to display fields - ALDRAN.PS000T01.RURAL.PQqFYyy
6. Rural Master Frame File reformatted to the same format as the Sample Selection File ALDRAN.PS400T01.REXTRACT.PQqFYyy

SELECTING THE FIRST STAGE SAMPLE

The first stage sample is a stratified random sample of route-days. There are four steps in this process. The first step is stratification, when routes with similar characteristics are grouped. In the second step, allocation, the number of routes to be sampled in each stratum is determined. In the third step, selection, routes from each stratum are randomly selected. In the fourth step, test dates are randomized and assigned to selected routes, thereby determining the route-days to be sampled. Each postal quarter (PQ), a new sample of route-days is selected, independently from those selected in prior quarters.

Stratification

Stratification is the process of assigning units with similar characteristics to the same group. Approximately four weeks prior to the start of a PQ, eligible rural carrier routes are stratified on the basis of the number of eligible routes within a district.

The two strata for the Rural Carrier Cost System are:

- 1) 1C for routes in districts with more than 20 routes in the district
- 2) 2C for routes in districts with less than 21 routes in the district

The program DDARAN.TESTSRC(ALD401C7) copies and reformats the rural frame file into two different files. The first is in the same structure as the rural frame but with the packed fields unpacked. The second has the master frame information but reformatted into the same structure as the sample selection file. DDARAN.TESTSRC(ALD403C5) stratifies the Rural Master Frame and adds counts for the two rural strata to the totals file.

INPUTS

1. Rural Master Frame: DSN = File - AAQMNN.PS700T01.RURLEXTR.PQqFYyy

OUTPUTS

1. Total number of routes in each stratum: DSN = &&CRTOTALS
(temporary SAS data set)
2. Rural Extract records eligible for sample selection: DSN = &&REXTRCT2
(temporary SAS data set)

The program ALDRAN.CCSREDES.PDS.SRC(ALD102C1) retrieves the base unit site code from the ZIPTABLE file and merges it into the Rural Master Frame file of eligible routes in preparation for sampling. The ZIPTABLE file is read and sorted in 3-digit ZIP Code order. The Rural Extract file is read and sorted by the first 3 digits of the Route ZIP Code (5-digit). The two files are then

merged by 3-digit ZIP Code, adding the site code field from the ZIPTABLE file to the reformatted Rural Master Frame. The new reformatted Master File is then sorted by site code, finance number, and route number and written to a temporary file as input to the sampling program.

INPUT

1. ZIPTABLE HSFRAN.PS001D01.ZIPTABLE.ORFEO (Maps 3 Digit Zip Code to site code)
2. Rural Extract from Program ALD100C1: DSN = &&REXTRCT2

OUTPUT

1. Rural Extract with the Site Code: DSN = &&REXTRCT3

Sample Allocation

The primary sampling unit is the route-day. The sampling percentages change slightly for the Rural Sample selection and the number of sample routes changes depending on the size of the Master Frame. The allocation does not vary much from quarter to quarter. However, the sample size is larger in the fourth quarter due to the increase in the number of delivery days. The sampling percentages are stored in the first row of the DATE file.

A copy of an actual percentage record on the date file is '0273601143023450104069A0228406854.' The first 20 digits are the sampling percentages for the City Carrier system. The next two digits are the number of delivery days in the quarter. Then comes a sort character used to insure that the percentage record is the first record in the date file. Following the sort character are the two Rural Sampling percentages. The first five digits are the percentage for stratum 1C and the remaining five are for stratum 2C.

Sample Selection

Each postal quarter, a systematic sample of route-days is randomly selected from each stratum. First, a random sample of routes is systematically selected without replacement. Then, randomized delivery dates are systematically assigned to selected routes. These steps are discussed in more detail below.

Based on sampling percentages, the sample size is determined for each stratum. The skip interval for each stratum is determined by dividing the stratum sample size into the number of routes for that stratum. As each route is selected for the sample, the test date is sequentially assigned from the randomized test date file. Test dates are assigned without replacement until all dates have been used, and then they are reused in the same randomized, sequential order.

A six-digit test identification number is assigned to each selected route-day, and is used for tracking tests throughout subsequent processing. The test identification number starts with the postal quarter number followed by a 4-digit sequential number and a one-digit check sum. The check sum is computed using the 'MODULUS 10' check digit algorithm, whereby the first, third, and fifth digits of the test identification are multiplied by two, and all digits of these products are added along with the second and fourth digits of the test identification number. The check sum is the difference between that number and the next highest multiple of 10.

For example, if the Test ID is 14627

1 st digit	2 nd digit	3 rd digit	4 th digit	5 th digit						
1*2=2		6*2=12		7*2=14						
		/		/						
		/		/						
2	+	4	+	1+2	+	2	+	1+4	=	16

The next highest multiple of 10, after 16, is 20, so the check sum is 20-16=4.

The complete Test Identification Number is 146274.

Program DDARAN.PRODSRC(ALD410C8) generates the random selection of rural routes from the extracted files created in Programs ALD102C1. The District Code is added to the sample record.

INPUTS

1. City Extract1 File: DSN = &&REXTRACT3
2. Group Totals - eligible routes for each stratum: DSN = &&CRTOTALS
3. Percentage And Dates: DSN = &&RANDATE
4. District codes: DSN = HSFRAN.PS001D01.ZIPTABLE.ORFEO

OUTPUT

1. Rural Samples - Random sample of rural routes: DSN= &&RSAMPLES

Program DDARAN.PRODSRC(ALD405C7) randomizes the DATES File by randomly numbering the delivery dates for a postal quarter. This ensures that the sample tests and test dates are randomly assigned.

INPUTS

1. Date File – ALDRAN.FY2000.PARMLIB(DATE.PQq)

OUTPUT

1. Percentage And Dates: DSN = &&RANDATE

Program DDARAN.PRODSRC(ALD412C5) assigns the test dates to the rural sample records. Test dates are assigned sequentially to the rural sample.

INPUTS

1. Rural Samples - Random sample of rural routes: DSN = &&RSAMPLES
2. Percentage And Dates: DSN = &&RANDATE

OUTPUT

1. Rural Sample - Random Rural sample selection with test dates: DSN = &&RSAMPLES

The sampled records are sorted into various sequences for generating reports.

Program ALD420C0 creates a file with one record for each district by selecting the first record of each District group for use in downstream processing.

INPUT

1. ZIPTABLE: DSN = HSFRAN.PS001D01.ZIPTABLE.ORFEO

OUTPUT

1. Consolidated District File - one record per District: DSN = ALDRAN.PS420D01.MSC

Program (ALD430C0) generates summary reports of number of tests by district, CAG, area, plant, and stratum.

INPUTS

1. Samples - city samples merged in a prior sort : DSN = &&SAMPLES1
2. District - consolidated list of district's: DSN = ALDRAN.PS420D01.MSC

OUTPUT Reports

1. ALD430P2 - summary of sample routes by District
2. ALD430P3 - summary of sample routes by CAG
3. ALD430P4 - summary of sample by area
4. ALD430P5 - summary of sample by plant
5. ALD450P3 - summary of sample routes by Stratum

Joint Scheduling

Most data collectors within a district conduct tests from many of the major statistical systems: ODIS, RPW, TRACS, City Carrier Cost, Rural Carrier Cost, etc. Selecting test dates for each system independently causes some days with substantially more tests than other days. To help alleviate this peak load problem, tests for ODIS, RPW, CCCS, and RCCS are jointly scheduled, taking into account the workload for TRACS tests. Although the sample selection programs for each of these systems randomly assign testdates, there are only three groups of tests for which those originally assigned testdates are retained: ODIS Sunday and holiday tests; RPW Sunday and holiday tests; and, TRACS tests.

The available dates for the postal quarter are randomized in a manner that ensures good temporal distribution. For a typical postal quarter with three accounting periods, any three consecutive records in the randomized list of dates will have one date from each accounting period, and any six consecutive records will have one date from each pay period.

Sample files for the five systems are concatenated, and Sunday and holiday tests, along with TRACS tests, are excluded from joint scheduling. Within a district, tests are sorted by ODIS

area (almost one-to-one correspondence with 3-digit ZIP Codes), size of test, and statistical system. Dates from the randomized date file are sequentially assigned to the ODIS, RPW, CCCS, and RCCS tests, taking into consideration the workload for TRACS tests. (NOTE: The Managers of Statistical Programs in each site are given some flexibility as to how tests are jointly scheduled in their sites.) For each system, the output sample file is the same format as the input sample files, and contains the same records, except the jointly scheduled test date and day of week have been substituted for the originals. After joint scheduling has been completed, the rural carrier sample is written out to the file ALDRAN.PS400T02.RURAL..PQqFYyy. (NOTE: The input rural carrier sample file is PS400T01).

Partitioning the Sample

The NOVA Statistical Programs Service Center (SPSC) is notified after CCCS, RCCS, RPW, TRACS, and ODIS tests have been jointly scheduled. They execute a C-list, which partitions the sample file for each base unit computer (typically, one per district). The sample files are then made available for each CODES site to download, after going through a two-step quality assurance process. In the first step, the NOVA SPSC downloads and checks the sample of one CODES site. After the first step is successfully completed, several CODES sites download and review the sample file. The sample file is made available to all CODES sites only after this second step is successfully completed.

INPUTS

1. Manual: Fiscal Year, Postal Quarter, Begin Quarter Date and End Quarter Date
2. Sample File: DSN=ALDRAN.PS400T02.RURAL.PQqFYyy

OUTPUTS

1. VSAM File – DSN=HSDHQV.VS340D01.CCS.SAMPLE

The VSAM file begins with an 11 byte key, where the first three (3) bytes are the Site ID, the next six (6) bytes are the Test ID, and the last two (2) bytes are blank for Rural records.

ADMINISTERING TESTS IN THE FIELD

Within each district, the Manager of Statistical Programs (MSP) is responsible for downloading the sample file, assigning tests to data collectors, reviewing the data after tests are conducted, and ensuring timely transmission of the data to the mainframe computer system at San Mateo.

Downloading the Sample

The CODES base unit contains a function that allows each site to download its sample. Site personnel utilize functions in the COMMUNICATION portion of the base unit software to download the sample. Once the site's sample is downloaded to its base unit, site personnel can download individual sample testid information to disks which allow a test to be performed on the CODES laptop computer.

Administrative Functions

Administrative tracking of tests is performed on the base unit computer(s) in each district. Administrative functions available on the base unit computer include the following: preparing disks with the sample file, so tests can be conducted; rescheduling tests; transferring data from tests taken on the CODES laptop computer to the base unit computer; deleting tests; tracking completion of tests; tracking transmission of tests to the mainframe; etc.

System documentation for these administrative functions is contained in USPS-LR-J-15, CCCS-RCCS CODES Computer System and Source Code, Section 4, Rural CCS CODES Base Unit Technical Documentation.

Collecting The Data

Data collectors knowledgeable in mail identification – determining the appropriate rate category of the mail and the special services provided - conduct the RCCS tests. The office in which the test is being conducted is called at least a day before the test to ensure that a carrier who knows the route will be available on the day of the test. In addition, the data collector speaks to the carrier supervisor and/or carrier to explain how the test will be conducted. The data collector typically arrives at the test facility as soon as the office is opened in order to identify and sample, with assistance from a supervisor, mail eligible for sampling. The data collector classifies each sampled mailpiece by class, subclass, and compensation category.

Data collection procedures are contained in Chapter 4 of Handbook F-65, Data Collection User's Guide for the Cost Systems (USPS-LR-J-14) with supplemental instructions in Supplemental Statistical Programs Policies and Data Collection Instructions (USPS-LR-J-34). The CODES software, used on laptop computers to record the data, is documented in USPS-LR-J-15, CCCS-RCCS CODES Computer System Documentation and Source Code, Section 3, Rural CCS CODES Data Entry Technical Documentation.

Transmitting the Data to San Mateo

Each week, the MSP transmits all tests from the base unit to the mainframe computer at San Mateo. A backup copy of the data is made as they are transmitted, in case of transmission errors or loss of data on the mainframe computer system.

DATA VALIDATION AND EDITING

Manual Editing

Each week the data are reviewed manually for errors and outliers. Data collectors or MSPs are generally contacted whenever an unusual number of pieces were recorded for a shape and subclass combination. When errors are identified, changes are recorded in an EXCEL file and later hard-coded into the edit programs to correct the data that were manually identified as errors or outliers.

Computerized Editing

A computerized edit is run each pay period to help ensure data quality. The test identification number (testid) is merged with a copy of the sample selection file to ensure that the testid is correct. The combination of shape and subclasses or rate category is validated to insure that each shape record identifies a valid mailpiece. The recorded numbers of pieces are checked against a list of high counts by shape/rate category to identify outliers.

The job that executes the edits is ALDRAN.FYyyyyQq.RURAL.CNTL(ALD300JP) The actual file is ALDRAN.FY2000Q1.RURAL.CNTL(ALD300JP). Several programs are executed within ALDRAN.FYyyyyqq.RURAL.CNTL(ALD300JP). The source code for all of these programs is in the file ALDRAN.FYyyyyQq.RURAL.PARMLIB.

Program ALD307J1 is a COBOL program that flags possible errors in the Route file.

Program ALD315J1 is a COBOL program that flags possible errors in the Shape file.

The output files are flat files that are used in the estimation process. The layouts for these files are the same as the layouts for the corresponding input files.

INPUTS

1. Route File – DSN=HSDHQN.PS345D01.CODES.RROUTE.CCSFYyyPPppw
2. Shape File – DSN=HSDHQN.PS345D01.CODES.RSHAPE.CCSFYyyPPppw

(Two weeks per pay period – week 1 and week 2 for each of the route and shape files)

1. Link mailcodes to subclass/shape – DSN=ALDRAN.FY2000.SORTED.MAILCODE(RURAL)
(LAYOUT 003)

OUTPUTS

1. Route File – DSN=ALDRAN.ROUTE.RCSFYyyPPpp (LAYOUT 001)
2. Shape File – DSN=ALDRAN.SHAPE.RCSFYyyPPpp (LAYOUT 002)

ESTIMATION AND VARIANCE

The RCCS produces two types of estimates—volumes and distribution keys (volume proportions). Estimates are generally computed on a quarterly and annual basis, and the annual volume estimates are the sum of the four quarterly estimates. This section provides the formulas used for FY 2000 to calculate the volumes, distribution keys, and the coefficients of variation associated with those estimates.

Notation:

y	variable of interest
w	weighting factor
h	postal quarter
i	compensation category domain
j	subclass or rate category domain
k	stratum
l	route-day
N	universe count – the number of routes in the stratum
n	completed tests in the stratum
d	delivery days in the postal quarter
s	skip utilized on a record being keyed by the data collector (third stage weight)
r	intermediate office weight (second stage weight)
\hat{Y}	estimate of the total
\hat{R}	estimate of the distribution key
Cov	estimate of the covariance
\hat{V}	estimate of the variance
CV	estimate of the coefficient of variation

The weight applied to each record consists of four parts. First is the number of delivery days, d , in each quarter. Second is the first stage weight, indicated by N_{hk}/n_{hk} . Third is the intermediate office weight r (second stage), which is applied on a testid basis and is almost always 1. Finally there is the skip interval, s , which is applied to each record in a test. Dividing by 1000 causes the estimates to be reported in thousands. This weighting process yields unbiased estimates of mail volumes assuming any missing tests are missing at random.

The weighting factor is:

$$w_{hk} = \left(\frac{d_h \times N_{hk} \times r \times s}{n_{hk} \times 1000} \right)$$

Variates are defined as follows:

$$y'_{hijkl} = \begin{cases} y_{hijkl} & \text{if the unit is in the } i^{\text{th}} \text{ and } j^{\text{th}} \text{ domains} \\ 0 & \text{otherwise} \end{cases}$$

$$x'_{hikl} = \begin{cases} x_{hikl} & \text{if the unit is in the } i^{\text{th}} \text{ domain} \\ 0 & \text{otherwise} \end{cases}$$

The quarterly volume for the intersection of the i^{th} compensation category domain and j^{th} subclass is

$$\hat{Y}_{hij} = \sum_k \sum_l w_{hk} y'_{hijkl}$$

The quarterly volume for the i^{th} compensation category domain is

$$\hat{X}_{hi} = \sum_k \sum_l w_{hk} x'_{hikl}$$

The quarterly distribution key for the i^{th} compensation category domain and j^{th} subclass is

$$\hat{R}_{hij} = \frac{\hat{Y}_{hij}}{\hat{X}_{hi}}$$

The annual volume for the i^{th} compensation category domain and j^{th} subclass is

$$\hat{Y}_{ij} = \sum_{h=1}^4 \hat{Y}_{hij}$$

The annual volume for the i^{th} compensation category domain is

$$\hat{X}_i = \sum_{h=1}^4 \hat{X}_{hi}$$

The annual distribution key for the intersection of the i^{th} and j^{th} domains is

$$\hat{R}_{ij} = \frac{\hat{Y}_{ij}}{\hat{X}_i}$$

Variance

In computing the sampling error on the estimates, an ultimate cluster variance estimator is used. An assumption is made that the sampling error within routes is very small relative to the overall sampling error. Therefore, the variance formula used is similar to a single-stage total or ratio estimate, except that it omits the finite population correction (fpc) factor.

The estimated stratum mean by postal quarter for the intersection of the i^{th} and j^{th} domains is

$$\bar{y}'_{hijk} = \frac{\sum_l y'_{hijkl}}{n_{hk}}$$

$$\hat{S}_{hijk}^2 = \frac{\sum_l (y'_{hijkl} - \bar{y}'_{hijk})^2}{n_{hk} - 1}$$

The estimated stratum variance for the quarterly volume for the intersection of the i^{th} and j^{th} domains is

$$V(\hat{Y}_{hijk}) = \frac{w_{hk}^2 \hat{S}_{hijk}^2}{n_{hk}}$$

The estimated variance for the quarterly volume for the intersection of the i^{th} and j^{th} domains is

$$V(\hat{Y}_{hij}) = \sum_k V(\hat{Y}_{hijk})$$

The estimated variance for the annual volume for the intersection of the i^{th} and j^{th} domains is

$$V(\hat{Y}_{ij}) = \sum_h V(\hat{Y}_{hij})$$

The estimated stratum mean by postal quarter for the intersection of the i^{th} domain is

$$\bar{x}'_{hikl} = \frac{\sum_l x'_{hikl}}{n_{hk}}$$

$$S_{hik}^2 = \frac{\sum_l (x'_{hikl} - \bar{x}'_{hik})^2}{n_{hk} - 1}$$

The estimated stratum variance for the quarterly volume for the i^{th} domain is

$$V(\hat{X}_{hik}) = \frac{w_{hk}^2 \hat{S}_{hik}^2}{n_{hk}}$$

The estimated variance for the quarterly volume for the i^{th} domain is

$$V(\hat{X}_{hi}) = \sum_k V(\hat{X}_{hik})$$

The estimated variance for the annual volume for the i^{th} domain is

$$V(\hat{X}_i) = \sum_h V(\hat{X}_{hi})$$

The estimated stratum covariance between the quarterly volumes for the intersection of the i^{th} and j^{th} domains is

$$Cov(\hat{Y}_{hijk}, \hat{X}_{hik}) = w_{hk}^2 \hat{S}_{yxbhijk}$$

where

$$\hat{S}_{yxbhijk} = \frac{\sum_l (y'_{hijkl} - \bar{y}'_{hijk})(x'_{hikl} - \bar{x}'_{hik})}{n_{hk} - 1}$$

The estimated covariance between the quarterly volumes for the intersection of the i^{th} and j^{th} domains is

$$Cov(\hat{Y}_{hij}, \hat{X}_{hi}) = \sum_k Cov(\hat{Y}_{hijk}, \hat{X}_{hik})$$

The estimated covariance between the annual volumes for the intersection of the i^{th} and j^{th} domains is

$$Cov(\hat{Y}_{ij}, \hat{X}_i) = \sum_h Cov(\hat{Y}_{hij}, \hat{X}_{hi})$$

The estimated relative variance (the square of the coefficient of variation) for the quarterly distribution key for the intersection of the i^{th} and j^{th} domain is

$$(CV)^2(\hat{R}_{hij}) = \left(\frac{V(\hat{Y}_{hij})}{\hat{Y}_{hij}^2} + \frac{V(\hat{X}_{hi})}{\hat{X}_{hi}^2} - \frac{2Cov(\hat{Y}_{hij}, \hat{X}_{hi})}{\hat{X}_{hi}\hat{Y}_{hij}} \right)$$

The estimated relative variance for the annual distribution key for the intersection of the i^{th} and j^{th} domain is

$$(CV)^2(\hat{R}_{ij}) = \left(\frac{V(\hat{Y}_{ij})}{\hat{Y}_{ij}^2} + \frac{V(\hat{X}_i)}{\hat{X}_i^2} - \frac{2Cov(\hat{Y}_{ij}, \hat{X}_i)}{\hat{X}_i\hat{Y}_{ij}} \right)$$

Quarterly Volume Estimates and Distribution Keys

Once the rural carrier data for an entire quarter has been edited, quarterly volume estimates and distribution keys are produced. The quarterly volumes are used as an input to the quarterly CRA and to review the estimates. The estimated volumes are compared with the same period the previous year and with estimates from other statistical systems. Substantial differences between the reports are investigated for additional quality assurance. The quarterly distribution keys programs are as follows:

Quarterly estimation is a four-step process. First, pay period files are concatenated to form the quarterly file. Second, collection mail volume estimates are calculated. Third, delivery volume estimates are calculated. Fourth, the Z file is produced.

ALDRAN.FY2000Qq.RURAL.CNTL(TWUKEYA9) is run to concatenate pay period files to form the quarterly file.

The input files are the edited pay (pp) period files.

INPUTS:

1. Edited Pay Period Data Files – DSN=ALDRAN.SHAPE.RCS00pp

Example for FY 00 PP20: ALDRAN.SHAPE.RCS0020

Pay periods included in each quarter are:

Only those tests that actually belong in the quarter (indicated by the first digit of the testid) are used for estimation. Below is a list of the pay periods that should be used as inputs for the estimation for each quarter.

PQ1 includes PP20, PP21, PP22, PP23, PP24, and PP25.

PQ2 includes PP26, PP01, PP02, PP03, PP04, and PP05.

PQ3 includes PP06, PP07, PP08, PP09, PP10, and PP11.

PQ4 includes PP12, PP13, PP14, PP15, PP16, PP17, PP18, and PP19.

The program is generally run more than once, so various global analyses and edits may be performed. Additionally, weights for the second and third stages of sampling are applied to the data.

OUTPUTS

The SAS dataset DSN = ALDRAN. RURAL.SASDSNS.SHAPE.FILE.FY2000Qq with SAS members RAWSH4 and TESTCNT. RAWSH4 contains all of the raw data records for the quarter and TESTCNT includes a listing of all test identification numbers for the quarter (used for weight development).

ALDRAN.FY2000Qq.RURAL.CNTL(WKEYA1) processes collection mail counts and produces first-stage weights to be applied to the data received from the TWUKEYA9 program. It executes two sets of SAS code. The first one DSN=ALDRAN.FY2000Qq.RURAL.PARMLIB(ALD750JZ) calculates the first-stage weights applied to all weighted volume estimates. The second one DSN=ALDRAN.FY2000Qq.RURAL.PARMLIB(ALD750X4) calculates the weighted volumes for collection mail data.

INPUTS:

Rural Master frame for universe counts DSN=AAQMNN.PS700T01.RURLEXTR.PQq00

Date file for number of delivery days in the quarter

DSN=ALDRAN.FY2000.PARMLIB(ATEPQq)

Data file for number of tests returned

DSN=ALDRAN.RURAL.SASDNS.SHAPE.FILE.FY2000Qq

Sample file for sample size DSN=ALDRAN.PS400T02.RURAL.PqqFY00

File containing edited collection tests DSN=ALDRAN.FY00.QqDET.EDIT

Flat file containing all previously calculated weights

DSN=ALDRAN.RURAL.WEIGHTS(FY2000Qq)

SAS file with weights to be used later DSN=ALDRAN.RCS2000.PQq.YTDWGT.DATA

OUTPUTS

SAS file with weights for processing data DSN=ALDRAN.RCS2000.PQq.YTDWGT.DATA

Quarterly collection volumes for Key Distribution

DSN=ALDRAN.LOTUS.RURAL.FY2000.PQq.DATA

ALDRAN.FY2000Qq.RURAL.CNTL(WKEYB1) processes delivered mail counts. It merges 1) the stratum from the sample selection file, 2) the weights for each stratum from the weights file, and 3) the mail category information from the mailcode file onto the raw mail counts file. The program then rolls up the information to two levels – mailcode, for external use, and CRA Bucket, for internal use.

INPUTS

File with weights DSN=ALDRAN.RCS2000.PQq.YTDWGT.DATA

File with mail category information for the mailcode output file

DSN=ALDRAN.FY2000Qq.SORTED.MAILCODE(RURAL)

File with mail category information for the CRA bucket output file

DSN=ALDRAN.SASAUTOS.RURMACRO.LIB2000(FORMATS)

Rural quarterly data file (SAS file) DSN=ALDRAN.RURAL.SASDNS.SHAPE.FILE.FY2000Qq
member RAWSH4

Sample file for strata DSN=ALDRAN.PS400T02.RURAL.PQqFY00

OUTPUTS

ALDRAN.FY00.RURAL.Qq.MCODE – Weighted data for each mailcode (Layout 005)

ALDRAN.FY00.RURAL.Qq.CRABKT – Weighted data for each CRA bucket.

ALDRAN.FY2000Qq.RURAL.CNTL(ZFILE) reproduces sections of the WKEYA9 and WKEYA1 programs to reproduce data by testid, mailcode, and skip. The resulting SAS data file ALDRAN.RURAL.Z.DATA.FY2000Qq is sent to the Postal Rate Commission. The SAS data set extension is RAWCNTS.

INPUTS

Edited Pay Period Data Files – DSN=ALDRAN.SHAPE.RCS00pp

Weight File – ALDRAN.RCS2000.PQq.YTDWGT.DATA

Sample Selection File – ALDRAN.PS400T02.RURAL.PQqFY00

Mailcode File – ALDRAN.FY2000Qq.SORTED.MAILCODE(RURAL)

OUTPUTS

Quarterly Z File – ALDRAN.RURAL.Z.FY2000Qq

Annual Estimates

Annual volumes are used to distribute costs to subclasses of mail. The volumes are calculated by summing the quarterly volumes. The annual volumes program is executed from the following file: ALDRAN.FY2000.RURAL.CNTL. Two members are utilized to produce the annual volumes.

ALDRAN.FY2000.RURAL.CNTL(COLL) is used to produce annual rural collection mail volumes.

INPUTS:

The quarterly volumes files are:

DSN=ALDRAN.LOTUS.RURAL.FY2000.PQ1.DATA

DSN=ALDRAN.LOTUS.RURAL.FY2000.PQ2.DATA

DSN=ALDRAN.LOTUS.RURAL.FY2000.PQ3.DATA

DSN=ALDRAN.LOTUS.RURAL.FY2000.PQ4.DATA

OUTPUT:

Annual volume report files: DSN=ALDRAN.LOTUS.RURAL.FY2000.COL.DATA – for collected mail volume estimates

ALDRAN.FY2000.RURAL.CNTL(SMIMCOD) is used to produce annual rural delivery mail volumes.

INPUTS:

The quarterly volumes files are:

DSN=ALDRAN.FY00.RURAL.Q1.MCODE

DSN=ALDRAN.FY00.RURAL.Q2.MCODE

DSN=ALDRAN.FY00.RURAL.Q3.MCODE

DSN=ALDRAN.FY00.RURAL.Q4.MCODE

OUTPUT:

The annual volume file is DSN=ALDRAN.LOTUS.RURAL.FY2000.FY00MC.DATA for delivered mail volume estimates.

FILE LAYOUTS

Rural Route (Header) File – Layout 001

<u>Position</u>	<u>Description</u>
1 - 6	Testid
7 - 10	Fiscal year in which test is taken
11 - 12	Month in which test is taken
13 - 14	Day on which test is taken
15 - 16	Pay period in which test is taken
17	Control Character
18 - 23	Finance Number
24	CAG
25 - 26	Roster designation
27	Test to be rescheduled
28	Rescheduled test
29 - 33	Route number
34 - 38	Route miles
39 - 40	Evaluated hours
41	Route Type
42	L-status
43	Number of intermediate routes
44	Route to be tested
45 - 49	ZIP Code of tested route
50 - 51	Number of hours DCT spent taking test
52 - 53	Number of minutes DCT spent taking test
54 - 56	DCT ID Number
57 - 61	Data entry software version
62 - 65	Base unit software version
66 - 73	Original date of test (if reschedule)
74 - 81	Filler
82 - 84	Site Number
85 -120	Filler

Rural Shape File – Layout 002

<u>Position</u>	<u>Description</u>
1 - 6	Testid
7 - 10	Fiscal year in which test is taken
11 - 12	Month in which test is taken
13 - 14	Day on which test is taken
15 - 16	Pay period in which test is taken
17	Session
18	Filler
19 - 20	Skip interval
21 - 23	Mailcode
24 - 26	Start
27 - 29	Mailcode
30 - 33	Data entry software version
34 - 37	Base unit software version
38 - 41	Sequence number of record
42	Intermediate route weight
43 - 50	Filler

Rurual Mailcode File Layout-003

Field	Position	Description
MAILCODE	1-6	Mailcode used for subclass/shape of mail being recorded
TYPE MAIL	7	Blank=non-accountable A=accountable
NAME	8-63	Description of Mailcode
BUCKET	64-65	Mail categories for Keys Program
N/A	66	
SHAPE CODE	67	A=letter B=flat C=boxholder D=parcel E=postage due F=Special services G=COD
BUCKET	68-69	See description below
DATAVER	71-74	Data Version Used
LOWER	76-78	The lowest possible piece count allowed without a flag.
UPPER	80-83	The highest possible piece count allowed without a flag.

Bucket	Description
1	First Class Mail (zero field)
2	First Class Single Piece
3	First Class Presort
4	First Class Single Piece Post Card
5	First Class Presort Post Cards
6	Total First Class (sum of 2-5)
7	Priority Mail
8	Express Mail
9	Mailgram
10	Periodicals
11	Standard Mail (A) (zero field)
12	Standard A Single Piece
13	Regular Standard (zero field)
14	Standard A Regular ECR
15	Standard A Regular Other
16	Standard A Total Regular (sum of 14-15)
17	Standard A Nonprofit (zero field)
18	Standard A Nonprofit ECR
19	Standard A Nonprofit Other
20	Standard A Total Nonprofit(sum of 18-19)
21	Total Standard A (sum of 12,16,and 20)
22	Standard Mail B (zero field)

Rurual Mailcode File Layout-003

23	<i>Standard Mail B Parcels Zone Rate</i>
25	<i>Standard B Special Standard</i>
26	<i>Standard B Library Rate</i>
24	<i>Standard B Bound Printed Matter</i>
27	<i>Total Standard B (sum of 23-26)</i>
28	<i>United States Postal Service</i>
29	<i>Free for the Blind and Handicapped</i>
30	<i>International Surface Mail (zero field)</i>
31	<i>International Surface Letter Class</i>
32	<i>International Surface AO</i>
33	<i>International Surface Parcel Post</i>
34	<i>International Surface M-Bag</i>
35	<i>International Airmail (zero field)</i>
36	<i>International Airmail Letter Class</i>
37	<i>International Airmail AO</i>
38	<i>International Airmail Parcel Post</i>
39	<i>International Airmail M-Bag</i>
40	<i>International Airmail Express</i>
41	<i>Total International Mail (sum of 31-34, and 36-40)</i>
42	<i>Total All Mail (sum of 6-10, 21, 27-29, and 41)</i>
43	<i>Special Services (zero field)</i>
44	<i>Accountable Mail Special Delivery</i>
45	<i>Not Used</i>
46	<i>Accountable Mail Certified</i>
47	<i>Accountable Mail COD</i>
48	<i>Accountable Mail Insured</i>
49	<i>Accountable Mail Registered</i>
50	<i>Accountable Mail Registered and COD</i>
51	<i>Accountable Mail Return Receipt</i>
52	<i>Accountable Mail Business Reply</i>
53	<i>Accountable Mail Other</i>
54	<i>Total Special Services (sum of 44-53)</i>

Rurual Mailcode File Layout-003

<i>55</i>	<i>Other</i>
<i>56</i>	<i>Bundled Mail</i>
<i>57</i>	<i>BSPS</i>

Rural Sample Selection File Layout-004

*** FORMAT DESCRIPTION ***										
UPDATED 6/3/97				DESCRIPTION RURAL CARRIER SAMPLES						
IDENT. (FLN)	DESCRIPTION	POS FROM	POS TO	LEN GTH	F O R M	J O R M	F I L E	D I C T I O N	OCCURS	C O M M E N T S
	FILLER	1	2	2	X		X			
	FINANCE NUMBER	3	8	6	X					
	SERVICE	9	9	1	X					
	AREA CODE	10	10	1	X					
	CAG	11	11	1	X					
	GROUP	12	13	2	X					'1C' OR '2C'
	P.O. NAME	14	43	30	X					
	STATE ABBREV.	44	45	2	X					
	ROUTE ZIP	46	50	5	X					
	ROUTE TYPE	51	54	4	X					
	ZIP-4-SUFFIX	55	58	4	X					
	FILLER	59	59	1	X		X			
	DATE OF TEST	60	67	8		9				
	DAY OF TEST	68	70	3	X					
	TESTID	71	76	6		9				
	L-STATUS	77	77	1	X					

*** FORMAT DESCRIPTION ***										
UPDATED 6/3/97				DESCRIPTION RURAL CARRIER SAMPLES						
IDENT. (FLN)	DESCRIPTION	POS FROM	POS TO	LEN GTH	F O R M	J O R M	F I L E	D I C T I O N	OCCURS	C O M M E N T S
	MILES	78	80	3		9				

Rural Sample Selection File Layout-004

STOPS	81	84	4	9					
FILLER	85	85	1	X	X				
ROUTE NUMBER	86	90	5	X					
CSD CODE	91	93	3	X					
TEST MONTH	94	96	3	X					
BOXES-REGULAR	97	101	5	9					
BOXES-CENTRAL	102	106	5	9					
BOX-FACTOR	107	110	4	9					
BOX-NUMBER	111	112	2						
DIVISION-ZIP-PRE	113	115	3						
FILLER	116	117	2	X	X				
SITE CODE	118	120	3						

Rural Mcode File – Layout 005

<u>Position</u>	<u>Description</u>
1 - 15	Volume
18 - 20	Mailcode
22 - 23	Bucket number
25 - 80	Mailcode description

Rural Mcode File – Layout 006

<u>SAS Variable Name</u>	<u>Description</u>
BKTCHAR	Letter Character
BKTNUM	Bucket Number
COMPLETE	Total number of completed delivery tests in the quarter
DELDAYS	Delivery days in the quarter
DELWGT	The first stage weight
F2848	Total number of collection mail forms completed in the quarter
INTWGT	Intermediate office weight (second stage weight)
MAILCODE	Mailcode for the record
MASTER	Stratum universe count of routes
NAME	Description of mailcode
SKIP	Skip interval for record (third stage weight)
STRATUM	Stratum in which the route (testid) exists
SUMPCS	Total unweighted mailpieces in a testid with the same skip mailcode
TESTID	Testid number for test
WGT	DELWGT/1000