

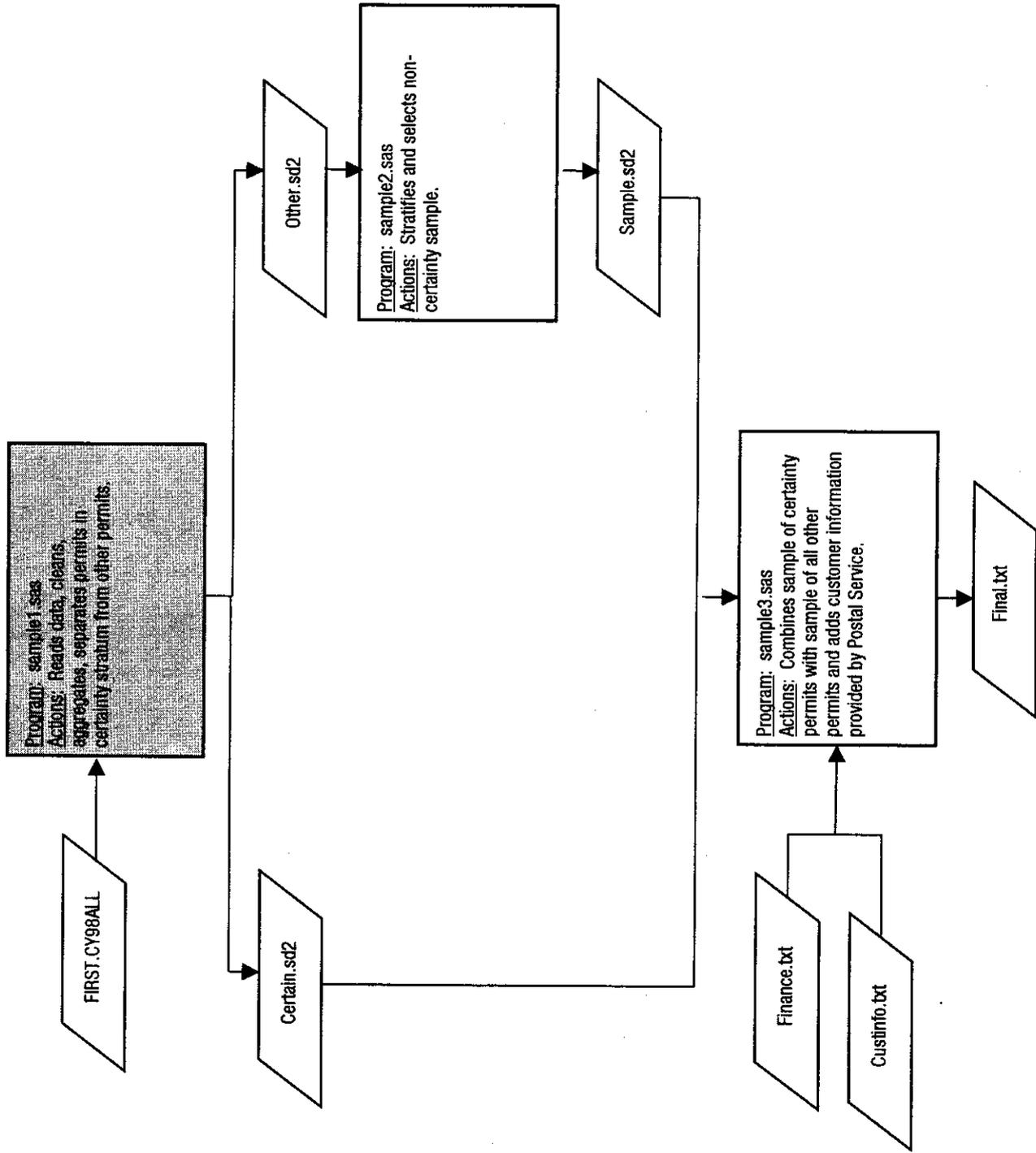
## Appendix B: Sample Design

### Contents

<b>Section</b>	<b>Page</b>
SECTION I: DOCUMENTATION OF SAMPLE1.SAS.....	1
SECTION II: DOCUMENTATION OF SAMPLE2.SAS.....	10
SECTION III: DOCUMENTATION OF SAMPLE3.SAS .....	21
SECTION IV: DOCUMENTATION OF SCREENER SAMPLE.SAS.....	32
SECTION V: DOCUMENTATION OF SCREENER SAMPLE_ADDITIONAL BATCH.SAS.....	49
SECTION VI: DOCUMENTATION OF QUESTIONNAIRE SAMPLE.SAS.....	64

## **Section I: Documentation of Sample1.sas**

# Flowchart of Sample Selection Programs



## FIRST-CLASS MAIL SAMPLE SELECTION PROGRAMS

**STUDY:** USPS Destination-Entry Discount Study  
**PROGRAM:** sample1.sas

### I. Requirements of Computer Analysis Relied Upon

#### A. General description of the program:

1. Objectives of the program: The objectives of this program are to read in the PERMIT data system extract file (FIRST.CY98ALL), aggregate volume and revenue variables, remove all non-bulk permits, and isolate permits that are part of the certainty stratum.
2. Processing tasks performed: This program reads raw data from FIRST.CY98ALL into a SAS data set, aggregates product volumes and revenues, removes permits that do not have any bulk volume, and separates out the permits to be included in the certainty stratum.
3. Methods and procedures employed: See attached program listing.
4. A listing of the input and output data: See attached pages describing names and sizes of input and output data.
5. A listing of source codes: See attached program and documentation.

#### B. Input data:

1. Designation of all sources of data: The input data set, FIRST.CY98ALL, is a database of all First-Class Mail permits from the PERMIT data system that were active in calendar year 1998. This data set was provided by the Postal Service.
2. Explanation of any modifications of the data made for use in the program: None.

**C. Definitions of all input and output variables or sets of variables:** See attached description of input and output data.

**D. A description of input and output data file organization:** See attached description of input and output data.

**E. A machine-readable copy of all databases:** The databases used in this program contain confidential data and are not provided.

**F. For all source codes, documentation sufficiently comprehensive and detailed to satisfy generally accepted software documentation standards appropriate to the type of program and to its intended use in the proceedings:** See attached program and documentation.

**G. The source program in machine-readable form:** The databases associated with this program are not provided due to confidentiality. The program can be provided upon request.  
Name of Program: sample1.sas

**H. All pertinent operating system and programming language manuals:** SAS Language: Reference, SAS Procedures Guide

- I. If requested program is interactive, a representative sample program run, together with any explanation necessary to illustrate the response sequence: N/A**
- J. "Canned" Statistical Packages: SAS v6.12 for Windows**
- K. Special requirements for computer simulations models offered if evidence or relied upon as support for other evidence: N/A**

**Program:** sample1.sas

**Input:**

(1) FIRST.CY98ALL

Description: FIRST.CY98ALL (CY98ALL) is the permit data file in text format provided by the Postal Service.

Number of observations: 61,976  
Number of variables: 43

**Outputs:**

(1) certain.sd2

Description: Certain.sd2 is the SAS data set containing permits which comprise the certainty stratum.

Number of observations: 41  
Number of variables: 13

(2) other.sd2

Description: Other.sd2 is the SAS data set containing all permits that were not included in the certainty stratum.

Number of observations: 40,848  
Number of variables: 13

**Description of variables in FIRST.CY98ALL:**

FIN_NUM	Finance number of issuing post office
INDICIA	Permit indicia type (Metered, Permit Imprint, Precanceled Stamp)
PERMIT	Permit number
CUST_ID	Customer identification number
CUSTOMER	Customer name
LCRAU_R	Automation carrier route letter revenue
CCRAU_R	Automation carrier route card revenue
L5DAU_R	Automation 5-digit letter revenue
C5DAU_R	Automation 5-digit card revenue
L3DAU_R	Automation 3-digit letter revenue
C3DAU_R	Automation 3-digit card revenue
F35AU_R	Automation 3/5-digit flat revenue
LBAAU_R	Automation basic letter revenue
FBAAU_R	Automation basic flat revenue
CBAAU_R	Automation basic card revenue
F35AUN_R	Automation 3/5-digit flat nonstandard revenue
FBAAUN_R	Automation basic flat nonstandard revenue
LPRNA_R	Nonautomation presort letter revenue
FPRNA_R	Nonautomation presort flat revenue
CPRNA_R	Nonautomation presort card revenue
PPRNA_R	Nonautomation presort parcel revenue
LPRNAN_R	Nonautomation presort letter nonstandard revenue
FPRNAN_R	Nonautomation presort flat nonstandard revenue
PPRNAN_R	Nonautomation presort parcel nonstandard revenue

LCRAU_V	Automation carrier route letter volume
CCRAU_V	Automation carrier route card volume
L5DAU_V	Automation 5-digit letter volume
C5DAU_V	Automation 5-digit card volume
L3DAU_V	Automation 3-digit letter volume
C3DAU_V	Automation 3-digit card volume
F35AU_V	Automation 3/5-digit flat volume
LBAAU_V	Automation basic letter volume
FBAAU_V	Automation basic flat volume
CBAAU_V	Automation basic card volume
F35AUN_V	Automation 3/5-digit flat nonstandard volume
FBAAUN_V	Automation basic flat nonstandard volume
LPRNA_V	Nonautomation presort letter volume
FPRNA_V	Nonautomation presort flat volume
CPRNA_V	Nonautomation presort card volume
PPRNA_V	Nonautomation presort parcel volume
LPRNAN_V	Nonautomation presort letter nonstandard volume
FPRNAN_V	Nonautomation presort flat nonstandard volume
PPRNAN_V	Nonautomation presort parcel nonstandard volume

**Description of new variables in certain.sd2 and other.sd2:**

NAUTO_R	Total nonautomation letter, flat, and parcel revenue
AUTO_R	Total automation letter, flat, and parcel revenue
CARD_R	Total card revenue
TOTAL_R	Total discounted First-Class Mail revenue
NAUTO_V	Total nonautomation letter, flat, and parcel volume
AUTO_V	Total automation letter, flat, and parcel volume
CARD_V	Total card volume
TOTAL_V	Total discounted First-Class Mail volume

**Actions of the Program:**

- Reads raw data from FIRST.CY98ALL.
- Aggregates detailed product volume and revenue into the necessary categories and drops all permits with no bulk volume.
- Sorts data by total bulk volume.
- Creates two SAS data sets containing certainty permits, certain.sd2, and non-certainty permits, other.sd2.

NOTE: Copyright (c) 1989-1996 by SAS Institute Inc., Cary, NC, USA.

NOTE: SAS (r) Proprietary Software Release 6.12 TS025  
Licensed to PRICE WATERHOUSE LLP, Site 0015509006.

```
1 /******  
2 /*  
3 /* PROGRAM: 'SAMPLE1.SAS' */  
4 /* INPUTS: 'FIRST.CY98ALL' - PERMIT DATA FILE IN TEXT FORMAT */  
5 /* TASKS: READ IN RAW DATA, AGGREGATE PRODUCT VOLUMES, IDENTIFY */  
6 /* LARGEST PERMITS TO BE INCLUDED IN CERTAINTY STRATUM */  
7 /* OUTPUTS: 'CERTAIN.SD2' - SAS DATASET WITH ONLY THE CERTAINTY PERMITS */  
8 /* 'OTHER.SD2' - SAS DATASET WITH ALL NON-CERTAINTY PERMITS */  
9 /* */  
10 /******  
11  
12 options nocenter nonumber nodate ls=120 ps=67;  
13 libname FCMDrop 'D:\FCMDROP';  
NOTE: Libref FCMDROP was successfully assigned as follows:  
Engine: V612  
Physical Name: D:\FCMDROP  
14 filename in1 'D:\FCMDROP\FIRST.CY98ALL';  
15  
16 /* Read in raw data file */  
17  
18 data permit;  
19 infile in1 lrecl=813 missover;  
20 input  
21 @1 fin_num $6. /* All variables are from PERMIT data system */  
22 @7 indicia $2. /* unless otherwise noted */  
23 @9 permit $5.  
24 @14 cust_id $10. /* from CBCIS customer table */  
25 @25 customer $51. /* from CBCIS customer table */  
26  
27 @86 LCRAU_R 13.2 /* Automation carrier route letter revenue */  
28 @99 CCRAU_R 13.2 /* Automation carrier route card revenue */  
29 @112 L5DAU_R 13.2 /* Automation 5-digit letter revenue */  
30 @125 C5DAU_R 13.2 /* Automation 5-digit card revenue */  
31 @138 L3DAU_R 13.2 /* Automation 3-digit letter revenue */  
32 @151 C3DAU_R 13.2 /* Automation 3-digit card revenue */  
33 @164 F35AU_R 13.2 /* Automation 3/5-digit flat revenue */  
34 @177 LBAAU_R 13.2 /* Automation basic letter revenue */  
35 @190 FBAAU_R 13.2 /* Automation basic flat revenue */  
36 @203 CBAAU_R 13.2 /* Automation basic card revenue */  
37 @229 F35AUN_R 13.2 /* Automation 3/5-digit flat nonstandard rev */  
38 @242 FBAAUN_R 13.2 /* Automation basic flat nonstandard revenue */  
39 @268 LPRNA_R 13.2 /* Nonautomation presort letter revenue */  
40 @281 FPRNA_R 13.2 /* Nonautomation presort flat revenue */  
41 @294 CPRNA_R 13.2 /* Nonautomation presort card revenue */  
42 @307 PPRNA_R 13.2 /* Nonautomation presort parcel revenue */  
43 @372 LPRNAN_R 13.2 /* Nonautomation presort letter nonstd rev */  
44 @385 FPRNAN_R 13.2 /* Nonautomation presort flat nonstd revenue */  
45 @398 PPRNAN_R 13.2 /* Nonautomation presort parcel nonstd rev */  
46  
47 @450 LCRAU_V 13.2 /* Naming for volume variables is identical to*/  
48 @463 CCRAU_V 13.2 /* revenue variables except all end in '_V' */  
49 @476 L5DAU_V 13.2  
50 @489 C5DAU_V 13.2  
51 @502 L3DAU_V 13.2  
52 @515 C3DAU_V 13.2  
53 @528 F35AU_V 13.2  
54 @541 LBAAU_V 13.2  
55 @554 FBAAU_V 13.2  
56 @567 CBAAU_V 13.2  
57 @593 F35AUN_V 13.2  
58 @606 FBAAUN_V 13.2  
59 @632 LPRNA_V 13.2  
60 @645 FPRNA_V 13.2  
61 @658 CPRNA_V 13.2  
62 @671 PPRNA_V 13.2  
63 @736 LPRNAN_V 13.2  
64 @749 FPRNAN_V 13.2  
65 @762 PPRNAN_V 13.2;  
66 run;
```

NOTE: The infile IN1 is:  
FILENAME=D:\FCMDROP\FIRST.CY98ALL,  
RECFM=V,LRECL=813

NOTE: 61976 records were read from the infile IN1.  
The minimum record length was 813.  
The maximum record length was 813.

NOTE: The data set WORK.PERMIT has 61976 observations and 43 variables.  
NOTE: The DATA statement used 33.6 seconds.

```
67
68 proc sort data=permit;
69   by customer;
70 run;
```

NOTE: The data set WORK.PERMIT has 61976 observations and 43 variables.  
NOTE: The PROCEDURE SORT used 58.1 seconds.

```
71
72 /* Aggregate volume and revenue data into totals for three categories */
73
74 data permit;
75 set permit;
76 NAUTO_R = LPRNA_R + FPRNA_R + PPRNA_R + /* Total non-auto L,F,&P revenue */
77   LPRNAN_R + FPRNAN_R + PPRNAN_R;
78 AUTO_R = LCRAU_R + L5DAU_R + L3DAU_R + /* Total auto L,F,&P revenue */
79   F35AU_R + LBAAU_R + FBAAU_R +
80   F35AUN_R + FBAAUN_R;
81 CARD_R = CCRAU_R + C5DAU_R + C3DAU_R + /* Total card revenue */
82   CBAAU_R + CPRNA_R;
83 TOTAL_R = NAUTO_R + AUTO_R + CARD_R;
84 NAUTO_V = LPRNA_V + FPRNA_V + PPRNA_V + /* Total non-auto L,F,&P volume */
85   LPRNAN_V + FPRNAN_V + PPRNAN_V;
86 AUTO_V = LCRAU_V + L5DAU_V + L3DAU_V + /* Total auto L,F,&P volume */
87   F35AU_V + LBAAU_V + FBAAU_V +
88   F35AUN_V + FBAAUN_V;
89 CARD_V = CCRAU_V + C5DAU_V + C3DAU_V + /* Total card volume */
90   CBAAU_V + CPRNA_V;
91 TOTAL_V = NAUTO_V + AUTO_V + CARD_V;
92 if TOTAL_V = 0 then delete; /* Delete permits with no bulk volume */
93 drop LCRAU_R CCRAU_R L5DAU_R C5DAU_R L3DAU_R C3DAU_R F35AU_R LBAAU_R FBAAU_R
94   CBAAU_R F35AUN_R FBAAUN_R LPRNA_R FPRNA_R CPRNA_R PPRNA_R
95   LPRNAN_R FPRNAN_R PPRNAN_R
96   LCRAU_V CCRAU_V L5DAU_V C5DAU_V L3DAU_V C3DAU_V F35AU_V LBAAU_V FBAAU_V
97   CBAAU_V F35AUN_V FBAAUN_V LPRNA_V FPRNA_V CPRNA_V PPRNA_V
98   LPRNAN_V FPRNAN_V PPRNAN_V;
99 run;
```

NOTE: The data set WORK.PERMIT has 40889 observations and 13 variables.  
NOTE: The DATA statement used 6.8 seconds.

```
100
101
102 /* Separate permits in the certainty stratum */
103
104 proc sort data=permit;
105   by descending TOTAL_V;
106 run;
```

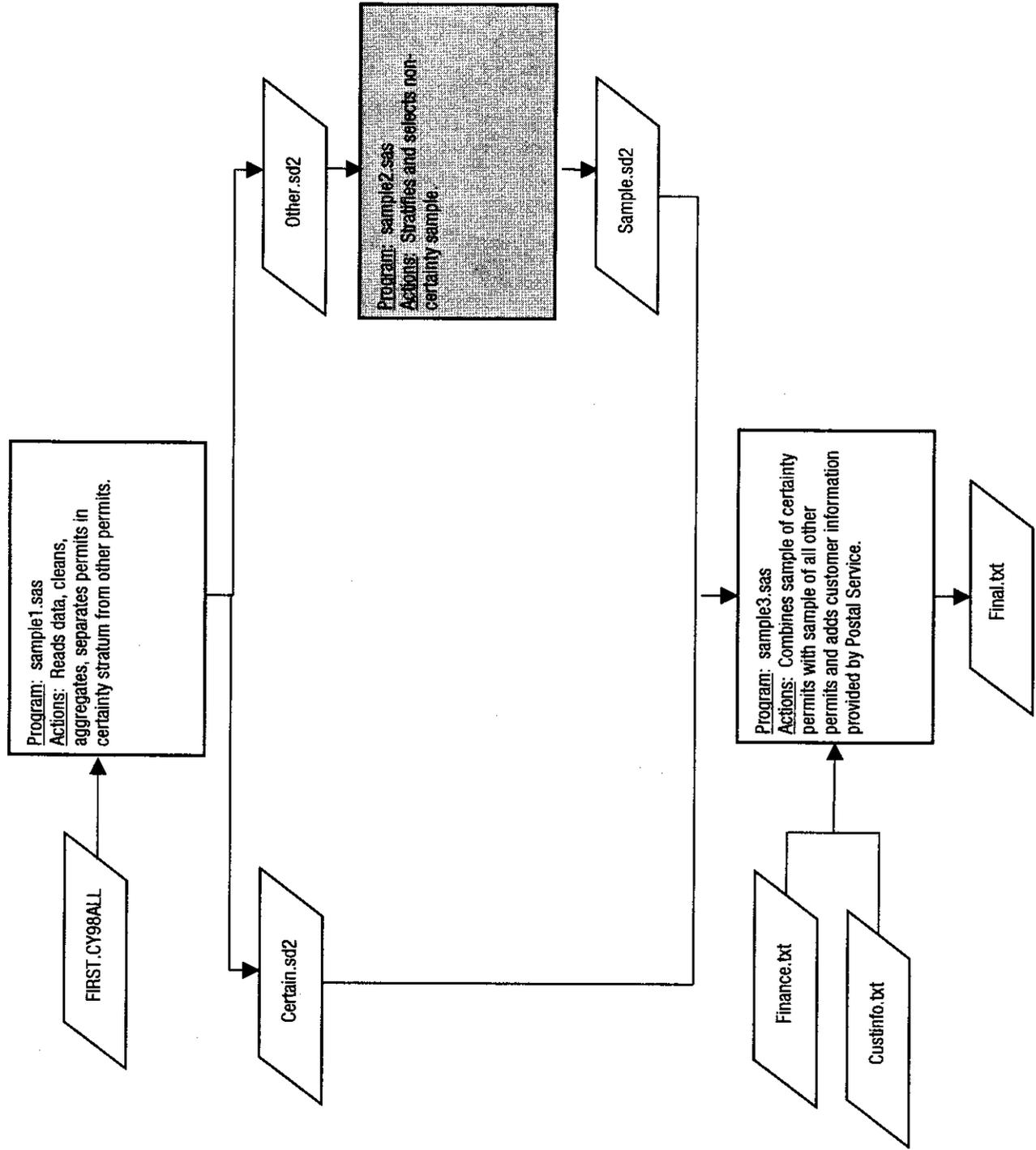
NOTE: The data set WORK.PERMIT has 40889 observations and 13 variables.  
NOTE: The PROCEDURE SORT used 2.75 seconds.

```
107
108 data FCMDrop.certain FCMDrop.other;
109 set permit;
110 if TOTAL_V => 130000000 then output FCMDrop.certain;
111 else output FCMDrop.other;
112 run;
```

NOTE: The data set FCMDROP.CERTAIN has 41 observations and 13 variables.  
NOTE: The data set FCMDROP.OTHER has 40848 observations and 13 variables.  
NOTE: The DATA statement used 2.75 seconds.

## **Section II: Documentation of Sample2.sas**

# Flowchart of Sample Selection Programs



## FIRST-CLASS MAIL SAMPLE SELECTION PROGRAMS

**STUDY:** USPS Destination-Entry Discount Study

**PROGRAM:** sample2.sas

### I. Requirements of Computer Analysis Relied Upon

#### A. General description of the program:

1. Objectives of the program: The objectives of this program are to divide the permits contained in the data set other.sd2 into three strata based on total discounted First-Class Mail volume and to randomly select a sample of permits within each stratum.
2. Processing tasks performed: This program selects stratum boundaries for the three strata that are formed from the permits contained in other.sd2, assigns each permit to a stratum, determines the appropriate stratum sample sizes, and randomly selects a sample of permits from each stratum.
3. Methods and procedures employed: See attached program listing.
4. A listing of the input and output data: See attached pages describing names and sizes of input and output data.
5. A listing of source codes: See attached program and documentation.

#### B. Input data:

1. Designation of all sources of data: The input data set is a SAS data set of non-certainty permits, other.sd2, from the program, sample1.sas.
2. Explanation of any modifications of the data made for use in the program: None.

**C. Definitions of all input and output variables or sets of variables:** See attached description of input and output data.

**D. A description of input and output data file organization:** See attached description of input and output data.

**E. A machine-readable copy of all databases:** The databases used in this program contain confidential data and are not provided.

**F. For all source codes, documentation sufficiently comprehensive and detailed to satisfy generally accepted software documentation standards appropriate to the type of program and to its intended use in the proceedings:** See attached program and documentation.

**G. The source program in machine-readable form:** The databases associated with this program are not provided due to confidentiality. The program can be provided upon request.

Name of Program: sample2.sas

**H. All pertinent operating system and programming language manuals:** SAS Language: Reference, SAS Procedures Guide

- I. If requested program is interactive, a representative sample program run, together with any explanation necessary to illustrate the response sequence: N/A**
- J. "Canned" Statistical Packages: SAS v6.12 for Windows**
- K. Special requirements for computer simulations models offered if evidence or relied upon as support for other evidence: N/A**

**Program:** sample2.sas

**Input:**

- (1) other.sd2  
Description: Other.sd2 is a SAS data set that was output from the program, sample1.sas. It contains all permits not included in the certainty stratum.
- |                         |        |
|-------------------------|--------|
| Number of observations: | 40,848 |
| Number of variables:    | 13     |

**Output:**

- (1) sample.sd2  
Description: Sample.sd2 is the SAS data set that was output from the program sample2.sas. It contains only permits selected as part of the sample from the non-certainty strata.
- |                         |     |
|-------------------------|-----|
| Number of observations: | 954 |
| Number of variables:    | 15  |

**Description of new variables:**

RAN_NUM	Random number used to select sample
STRATA	Number of the stratum that was assigned to each permit

**Actions of the program:**

- The Dalenius and Hodges method is used to determine the optimal stratum boundaries.<sup>1</sup>
- Neyman allocation is used to determine the optimal percentage of the sample to select from each stratum<sup>2</sup>
- Randomly selects permits in each stratum.

---

<sup>1</sup> Cochran, *Sampling Techniques*, John Wiley & Sons, 1977, pp. 127-130.

<sup>2</sup> Ibid., pp. 98-99.

NOTE: Copyright (c) 1989-1996 by SAS Institute Inc., Cary, NC, USA.

NOTE: SAS (r) Proprietary Software Release 6.12 TS025

Licensed to PRICE WATERHOUSE LLP. Site 0015509006.

```
1 /*****  
2 /*  
3 /* PROGRAM: 'SAMPLE2.SAS'  
4 /* INPUTS: 'OTHER.SD2' - OUTPUT FROM 'SAMPLE1.SAS' CONTAINS ALL PERMITS */  
5 /* NOT INCLUDED IN THE CERTAINTY STRATUM */  
6 /* TASKS: STRATIFY AND SELECT RANDOM SAMPLE FROM NON-CERTAINTY PERMITS */  
7 /* OUTPUTS: 'SAMPLE.SD2' - ALL PERMITS TO BE SAMPLED FROM THE NON- */  
8 /* CERTAINTY STRATA */  
9 /*  
10 /*****  
11  
12 options nocenter nonumber nodate ls=120 ps=67;  
13 libname FCMdrop 'D:\FCMDROP';  
NOTE: Libref FCMDROP was successfully assigned as follows:  
Engine: V612  
Physical Name: D:\FCMDROP  
14  
15 /* Use Dalenius-Hodges method to determine strata breakpoints */  
16  
17 /* Assign each non-certainty permit to a group defined by volume */  
18 /* increments of 500,000 pieces */  
19  
20 data other;  
21 set FCMdrop.other;  
22 lb = 0;  
23 ub = 500000;  
24 do i = 1 to 260;  
25 if TOTAL_V > lb and TOTAL_V <= ub then do;  
26 group = i;  
27 lower = lb;  
28 upper = ub;  
29 end;  
30 lb = lb + 500000;  
31 ub = ub + 500000;  
32 end;  
33 drop i lb ub;  
34 run;
```

NOTE: The data set WORK.OTHER has 40848 observations and 16 variables.

NOTE: The DATA statement used 7.95 seconds.

```
35  
36 proc sort data=other;  
37 by group;  
38 run;
```

NOTE: The data set WORK.OTHER has 40848 observations and 16 variables.

NOTE: The PROCEDURE SORT used 3.41 seconds.

```
39  
40 /* Calculate the number of permits in each volume group */  
41  
42 proc means data=other noprint;  
43 var TOTAL_V;  
44 id lower upper;  
45 by group;  
46 output out=groups(drop=_type_) sum=bulk_vol;  
47 run;
```

NOTE: The data set WORK.GROUPS has 183 observations and 5 variables.

NOTE: The PROCEDURE MEANS used 0.54 seconds.

```
48  
49 /* Calculate square root of frequency for each group and cumulative square root */  
50
```

```

51 data groups;
52 set groups;
53 if _N_ = 1 then cum_sr = 0;
54 sr_freq = sqrt(_freq_);
55 cum_sr = cum_sr + sr_freq;
56 retain cum_sr;
57 run;

```

NOTE: The data set WORK.GROUPS has 183 observations and 7 variables.  
NOTE: The DATA statement used 0.38 seconds.

```

58
59 /* Calculate total cumulative square root for all groups */
60
61 proc means data=groups noprint;
62 var sr_freq;
63 output out=tot_sr(drop=_type_ _freq_) sum=tot_sr;
64 run;

```

NOTE: The data set WORK.TOT\_SR has 1 observations and 1 variables.  
NOTE: The PROCEDURE MEANS used 0.05 seconds.

```

65
66 /* Add total cumulative square root variable onto each group record */
67
68 data groups FCMdrop.groups;
69 set groups;
70 if _N_ = 1 then set tot_sr;
71 run;

```

NOTE: The data set WORK.GROUPS has 183 observations and 8 variables.  
NOTE: The data set FCMDROP.GROUPS has 183 observations and 8 variables.  
NOTE: The DATA statement used 0.22 seconds.

```

72
73 /* Determine strata breakpoints based on the total cumulative square root */
74
75 /* Calculate the difference between the breakpoint square root values and */
76 /* the cumulative square root for each group */
77
78 data groups;
79 set groups;
80 array bp(3) bp1-bp3;
81 array diff(3) diff1-diff3;
82 do i = 1 to 3;
83 bp(i) = tot_sr / 3 * i;
84 diff(i) = cum_sr - bp(i);
85 if diff(i) < 0 then diff(i) = -1 * diff(i);
86 end;
87 drop i bp1-bp3;
88 run;

```

NOTE: The data set WORK.GROUPS has 183 observations and 11 variables.  
NOTE: The DATA statement used 0.55 seconds.

```

89
90 /* Use a macro to determine the breakpoint for each strata */
91 /* based on the group with a cumulative square root that is */
92 /* closest to the breakpoint square root values */
93
94 /* The macro will output one dataset for each breakpoint */
95
96 %macro break(bp, bp_var, out_set);
97
98 proc sort data=groups;
99 by &bp_var;
100 run;
101
102 data &out_set;
103 set groups;

```

```
104 strata = &bp;
105 if _N_ = 1 then output;
106 keep group strata lower upper;
107 run;
108
109 %mend break;
110
111 %break(1, diff1, bpoints);
```

NOTE: The data set WORK.GROUPS has 183 observations and 11 variables.  
NOTE: The PROCEDURE SORT used 0.11 seconds.

NOTE: The data set WORK.BPOINTS has 1 observations and 4 variables.  
NOTE: The DATA statement used 0.05 seconds.

```
112 %break(2, diff2, bp2);
```

NOTE: The data set WORK.GROUPS has 183 observations and 11 variables.  
NOTE: The PROCEDURE SORT used 0.05 seconds.

NOTE: The data set WORK.BP2 has 1 observations and 4 variables.  
NOTE: The DATA statement used 0.05 seconds.

```
113 %break(3, diff3, bp3);
```

NOTE: The data set WORK.GROUPS has 183 observations and 11 variables.  
NOTE: The PROCEDURE SORT used 0.05 seconds.

NOTE: The data set WORK.BP3 has 1 observations and 4 variables.  
NOTE: The DATA statement used 0.05 seconds.

```
114
115 /* Combine breakpoint datasets into one dataset with all breakpoints */
116
117 proc append base=bpoints data=bp2;
118 run;
```

NOTE: Appending WORK.BP2 to WORK.BPOINTS.  
NOTE: 1 observations added.  
NOTE: The data set WORK.BPOINTS has 2 observations and 4 variables.  
NOTE: The PROCEDURE APPEND used 0.28 seconds.

```
119
120 proc append base=bpoints data=bp3;
121 run;
```

NOTE: Appending WORK.BP3 to WORK.BPOINTS.  
NOTE: 1 observations added.  
NOTE: The data set WORK.BPOINTS has 3 observations and 4 variables.  
NOTE: The PROCEDURE APPEND used 0.05 seconds.

```
122
123 data bpoints;
124 set bpoints;
125 if _N_ = 1 then do;
126 temp = upper;
127 lower = 0;
128 end;
129 else lower = temp;
130 temp = upper;
131 retain temp;
132 drop temp;
133 run;
```

NOTE: The data set WORK.BPOINTS has 3 observations and 4 variables.  
NOTE: The DATA statement used 0.11 seconds.

```
134
135 /* Convert breakpoint dataset into one observation with three */
136 /* variables */
137
138 data strata;
139 set bpoints;
140 array bound(3) bound1-bound3;
141 do i = 1 to 3;
142   if i = strata then bound(i) = upper;
143 end;
144 retain bound1-bound3;
145 if _N_ < 3 then delete;
146 drop group lower upper strata i;
147 run;
```

NOTE: The data set WORK.STRATA has 1 observations and 3 variables.  
NOTE: The DATA statement used 0.11 seconds.

```
148
149 /* Add breakpoint variables to the 'OTHER' dataset */
150 /* Determine the appropriate stratum for each permit */
151
152 data other problem;
153 set other;
154 if _N_ = 1 then set strata;
155 if TOTAL_V <= bound1 then strata = 1;
156 if TOTAL_V > bound1 and TOTAL_V <= bound2 then strata=2;
157 if TOTAL_V > bound2 and TOTAL_V <= bound3 then strata=3;
158 if strata = . then output problem;
159 output other;
160 drop bound1-bound3;
161 run;
```

NOTE: The data set WORK.OTHER has 40848 observations and 17 variables.  
NOTE: The data set WORK.PROBLEM has 0 observations and 17 variables.  
NOTE: The DATA statement used 3.12 seconds.

```
162
163 proc sort data=other;
164   by strata;
165 run;
```

NOTE: The data set WORK.OTHER has 40848 observations and 17 variables.  
NOTE: The PROCEDURE SORT used 3.41 seconds.

```
166
167 /* Use Neyman allocation to determine sample size from each stratum */
168
169 /* Calculate the standard deviation of total bulk volume for each stratum */
170
171 proc means data=other noprint;
172   var TOTAL_V;
173   by strata;
174   output out=totals(drop=_type_) sum=std=std_vol;
175 run;
```

NOTE: The data set WORK.TOTALS has 3 observations and 4 variables.  
NOTE: The PROCEDURE MEANS used 0.33 seconds.

```
176
177 /* Calculate the stratum weights */
178
179 data totals;
180 set totals;
181 weight = _freq_ * std_vol;
```

182 run;

NOTE: The data set WORK.TOTALS has 3 observations and 5 variables.  
NOTE: The DATA statement used 0.05 seconds.

```
183
184 /* Calculte the total population weight */
185
186 proc means data=totals noprint;
187   var weight;
188   output out=temp(drop=_type_ _freq_) sum=tot_wt;
189 run;
```

NOTE: The data set WORK.TEMP has 1 observations and 1 variables.  
NOTE: The PROCEDURE MEANS used 0.05 seconds.

```
190
191 /* Merge the total weight back on to each stratum observation */
192
193 data totals;
194   set totals;
195   if _N_ = 1 then set temp;
196 run;
```

NOTE: The data set WORK.TOTALS has 3 observations and 6 variables.  
NOTE: The DATA statement used 0.05 seconds.

```
197
198 /* Calculate sample percentage (pi) for each stratum */
199
200 /* Use sample percentages to determine how many complete interviews */
201 /* with a target for total complete interviews of 200 */
202
203 /* Increase targets by an inflation factor of six */
204
205 data totals;
206   set totals;
207   pi = weight / tot_wt;
208   complete = round((200 - 41) * pi, 1); /* Target completes = 200 */
209   sample = round(6 * complete, 1); /* Inflation factor = 6 */
210 run;
```

NOTE: The data set WORK.TOTALS has 3 observations and 9 variables.  
NOTE: The DATA statement used 0.11 seconds.

```
211
212 /* Merge breakpoints on to strata dataset with sample sizes by strata */
213
214 data bpoints;
215   set bpoints;
216   bound = upper;
217   keep strata bound;
218 run;
```

NOTE: The data set WORK.BPOINTS has 3 observations and 2 variables.  
NOTE: The DATA statement used 0.05 seconds.

```
219
220 data totals;
221   merge totals bpoints;
222   by strata;
223 run;
```

NOTE: The data set WORK.TOTALS has 3 observations and 10 variables.  
NOTE: The DATA statement used 0.05 seconds.

```
224
225 /* Print sample size calculations */
```

```

226
227 proc print data=totals;
228   title 'Sample sizes by strata';
229   var strata _freq_ TOTAL_V weight pi complete sample bound;
230   sum pi weight complete sample;
231 run;

```

NOTE: The PROCEDURE PRINT used 0.48 seconds.

```

232
233 /* Assign a different random number to the permits in each stratum */
234
235 data other;
236   set other;
237   if strata = 1 then ran_num = ranuni(41303);
238   if strata = 2 then ran_num = ranuni(83609);
239   if strata = 3 then ran_num = ranuni(780);
240 run;

```

NOTE: The data set WORK.OTHER has 40848 observations and 18 variables.  
NOTE: The DATA statement used 3.02 seconds.

```

241
242 /* Sort by the random number */
243
244 proc sort data=other;
245   by strata descending ran_num;
246 run;

```

NOTE: The data set WORK.OTHER has 40848 observations and 18 variables.  
NOTE: The PROCEDURE SORT used 4.33 seconds.

```

247
248 /* Create a new dataset for each stratum */
249
250 data strata_1 strata_2 strata_3;
251   set other;
252   if strata = 1 then output strata_1;
253   if strata = 2 then output strata_2;
254   if strata = 3 then output strata_3;
255 run;

```

NOTE: The data set WORK.STRATA\_1 has 37739 observations and 18 variables.  
NOTE: The data set WORK.STRATA\_2 has 2688 observations and 18 variables.  
NOTE: The data set WORK.STRATA\_3 has 421 observations and 18 variables.  
NOTE: The DATA statement used 3.06 seconds.

```

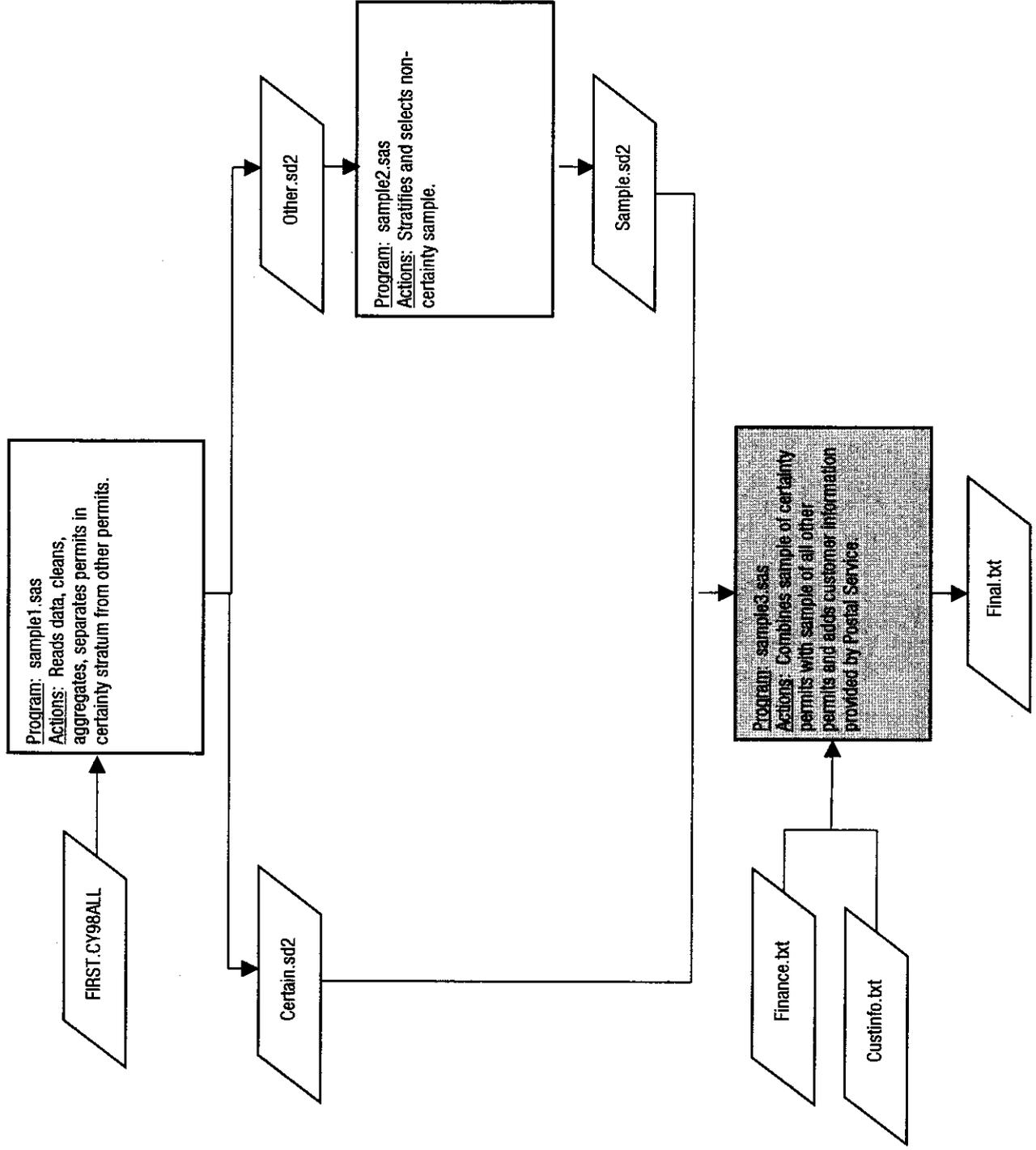
256
257 /* Use random number to select the appropriate number of permits from each stratum */
258 /* Create single dataset for all sampled permits */
259
260 data FCMdrop.sample;
261   set strata_1(obs=216) strata_2(obs=342) strata_3(obs=396);
262   drop group lower upper;
263 run;

```

NOTE: The data set FCMDROP.SAMPLE has 954 observations and 15 variables.  
NOTE: The DATA statement used 0.38 seconds.

## **Section III: Documentation of Sample3.sas**

# Flowchart of Sample Selection Programs



## FIRST-CLASS MAIL SAMPLE SELECTION PROGRAMS

**STUDY:** USPS Destination-Entry Discount Study

**PROGRAM:** sample3.sas

### I. Requirements of Computer Analysis Relied Upon

#### A. General description of the program:

1. Objectives of the program: The objectives of this program are to combine the permits contained in certain.sd2 with the sampled permits from sample.sd2 and to add additional information to each permit regarding the issuing post office and the permit owner.
2. Processing tasks performed: This program combines the sampled permits from certain.sd2 and sample.sd2 to form one SAS dataset, it merges the sampled permits against finance.txt to add information regarding the issuing post office, it merges the sampled permits against custinfo.txt to add information regarding the permit owner, and it outputs the sampled permits and all relevant information to a text file, final.txt.
3. Methods and procedures employed: See attached program listing.
4. A listing of the input and output data: See attached pages describing names and sizes of input and output data.
5. A listing of source codes: See attached program and documentation.

#### B. Input data:

1. Designation of all sources of data: Certain.sd2 is the output of program sample1.sas; sample.sd2 is the output of the program sample2.sas; financw.txt is a text file containing finance numbers and associated information (city, state, ZIP Code) that was provided by the Postal Service; and custinfo.txt is a text file containing customer ID numbers and associated information (address, contact, telephone number) that was provided by the Postal Service.
2. Explanation of any modifications of the data made for use in the program: None.

**C. Definitions of all input and output variables or sets of variables:** See attached description of input and output data.

**D. A description of input and output data file organization:** See attached description of input and output data.

**E. A machine-readable copy of all databases:** The databases used in this program contain confidential data and are not provided.

**F. For all source codes, documentation sufficiently comprehensive and detailed to satisfy generally accepted software documentation standards appropriate to the type of program and to its intended use in the proceedings:** See attached program and documentation.

**G. The source program in machine-readable form:** The databases associated with this program are not provided due to confidentiality. The program can be provided upon request.

Name of Program: sample3.sas

**H. All pertinent operating system and programming language manuals:** SAS Language: Reference, SAS Procedures Guide

**I. If requested program is interactive, a representative sample program run, together with any explanation necessary to illustrate the response sequence:** N/A

**J. "Canned" Statistical Packages:** SAS v6.12 for Windows

**K. Special requirements for computer simulations models offered if evidence or relied upon as support for other evidence:** N/A

**Program:** sample3.sas

**Inputs:**

- (1) certain.sd2  
Description: Certain.sd2 contains all permits in the certainty stratum from the program, sample1.sas.  

Number of observations:	41
Number of variables:	13
  
- (2) sample.sd2  
Description: Sample.sd2 is a SAS data set that was output from the program, sample2.sas. It contains all sampled permits that are not part of the certainty stratum.  

Number of observations:	954
Number of variables:	15
  
- (3) finance.txt  
Description: Finance.txt is a text file containing finance numbers and associated information (city, state, ZIP Code, etc.).  

Number of observations:	28,595
Number of variables:	5
  
- (4) custinfo.txt  
Description: Custinfo.txt is a text file containing customer IDs and associated information (address, city, state, ZIP Code, contact, etc.).  

Number of observations:	52,777
Number of variables:	8

**Output:**

- (1) final.txt  
Description: Final.txt is a text file containing all sampled permits and associated relevant information.  

Number of observations:	995
Number of variables:	26

**Description of variables in finance.txt:**

FIN_NUM	Postal Service finance number
POCITY	Post office city
POSTATE	Post office state
POZIP	Post office ZIP Code (5-digit)
POZIP4	Post office ZIP+4

**Description of variables in custinfo.txt:**

CUST_ID	Postal Service customer ID number
CUST	Customer name
ADDRESS	Customer street address
CITY	Customer city
STATE	Customer state
ZIP	Customer ZIP Code
CONTACT	Customer contact name
PHONE	Customer contact phone number

**Description of new variables in final.txt:**

NAUTO_U	Average revenue per piece for nonautomation volume
AUTO_U	Average revenue per piece for automation volume
CARD_U	Average revenue per piece for card volume

NOTE: Copyright (c) 1989-1996 by SAS Institute Inc., Cary, NC, USA.  
NOTE: SAS (r) Proprietary Software Release 6.12 TS025  
Licensed to PRICE WATERHOUSE LLP, Site 0015509006.

```
1 /*****  
2 /*  
3 /* PROGRAM: 'SAMPLE3.SAS'  
4 /* INPUTS: 'FINANCE.TXT' - TEXT FILE CONTAINING ALL FINANCE NUMBERS AND /*  
5 /* ASSOCIATED INFORMATION (CITY, STATE, ZIP, ETC.) /*  
6 /* 'CUSTINFO.TXT' - TEXT FILE CONTAINING CUSTOMER IDs AND /*  
7 /* ASSOCIATED INFORMATION (ADDRESS, CITY, STATE, ZIP, etc.) /*  
8 /* 'CERTAIN.SD2' - CERTAINTY STRATUM PERMITS FROM 'SAMPLE1.SAS' /*  
9 /* 'SAMPLE.SD2' - SAMPLED NON-CERTAINTY PERMITS FROM 'SAMPLE2.SAS' /*  
10 /* TASKS: COMBINE CERTAINTY PERMITS WITH OTHER SAMPLED PERMITS INTO ONE /*  
11 /* DATASET, MERGE ISSUING POST OFFICE INFORMATION ONTO SAMPLE BY /*  
12 /* FINANCE NUMBER, MERGE CUSTOMER INFORMATION ONTO SAMPLE BY /*  
13 /* CUSTOMER ID, OUTPUT SAMPLED PERMITS TO TEXT FILE /*  
14 /* OUTPUTS: 'FINAL.TXT' - TEXT FILE CONTAINING ALL SAMPLED PERMITS AND /*  
15 /* ASSOCIATED RELEVANT INFORMATION /*  
16 /*  
17 /*****  
18  
19 options nocenter nonumber nodate ls=120 ps=67;  
20 libname FCMdrop 'D:\FCMDROP';  
NOTE: Libref FCMDROP was successfully assigned as follows:  
Engine: V612  
Physical Name: D:\FCMDROP  
21 filename in1 'D:\FCMDROP\FINANCE.TXT';  
22 filename in2 'D:\FCMDROP\CUSTINFO.TXT';  
23 filename out1 'D:\FCMDROP\FINAL.TXT';  
24  
25 /* Assign a strat number of 4 to each permit in the certainty stratum /*  
26  
27 data certain;  
28 set FCMdrop.certain;  
29 strata = 4;  
30 run;
```

NOTE: The data set WORK.CERTAIN has 41 observations and 14 variables.  
NOTE: The DATA statement used 0.88 seconds.

```
31  
32 /* Combine permits in the certainty stratum with sampled permits from the /*  
33 /* other three strata /*  
34  
35 data sample;  
36 set FCMdrop.sample;  
37 run;
```

NOTE: The data set WORK.SAMPLE has 954 observations and 15 variables.  
NOTE: The DATA statement used 0.11 seconds.

```
38  
39 proc append base=sample data=certain;  
40 run;
```

NOTE: Appending WORK.CERTAIN to WORK.SAMPLE.  
WARNING: Variable RAN\_NUM was not found on DATA file.  
NOTE: 41 observations added.  
NOTE: The data set WORK.SAMPLE has 995 observations and 15 variables.  
NOTE: The PROCEDURE APPEND used 0.38 seconds.

```
41  
42 proc sort data=sample;  
43 by strata;  
44 run;
```

NOTE: The data set WORK.SAMPLE has 995 observations and 15 variables.  
NOTE: The PROCEDURE SORT used 0.22 seconds.

```
45
46 /* Read in the text file with finance numbers and associated information */
47
48 data finance;
49 infile in1;
50 input
51   @1 fin_num $6.
52   @7 POcity $22.
53   @30 POstate $2.
54   @66 POzip $5.
55   @71 POzip4 $4.;
56 run;
```

NOTE: The infile IN1 is:  
FILENAME=D:\FCMDROP\FINANCE.TXT,  
RECFM=V,LRECL=256

NOTE: 28595 records were read from the infile IN1.  
The minimum record length was 80.  
The maximum record length was 103.  
NOTE: The data set WORK.FINANCE has 28595 observations and 5 variables.  
NOTE: The DATA statement used 1.69 seconds.

```
57
58 /* Read in text file with customer IDs and associated information */
59
60 data custinfo;
61 infile in2 lrecl=149 missover;
62 input
63   @1 cust_id $10.
64   @12 cust $50.
65   @63 address $30.
66   @94 city $20.
67   @115 state $2.
68   @118 zip $5.
69   @124 contact $15.
70   @140 phone $10.;
71 run;
```

NOTE: The infile IN2 is:  
FILENAME=D:\FCMDROP\CUSTINFO.TXT,  
RECFM=V,LRECL=149

NOTE: 52777 records were read from the infile IN2.  
The minimum record length was 10.  
The maximum record length was 149.  
NOTE: The data set WORK.CUSTINFO has 52777 observations and 8 variables.  
NOTE: The DATA statement used 5.49 seconds.

```
72
73 /* Preserve sort order of all sampled permits */
74
75 proc sort data=finance;
76   by fin_num;
77 run;
```

NOTE: The data set WORK.FINANCE has 28595 observations and 5 variables.  
NOTE: The PROCEDURE SORT used 0.66 seconds.

```
78
79 data sample;
80 set sample;
81 sort_num = _N_;
82 run;
```

NOTE: The data set WORK.SAMPLE has 995 observations and 16 variables.  
NOTE: The DATA statement used 0.11 seconds.

```
83
84 proc sort data=sample;
85   by fin_num;
86 run;
```

NOTE: The data set WORK.SAMPLE has 995 observations and 16 variables.  
NOTE: The PROCEDURE SORT used 0.11 seconds.

```
87
88 /* Merge sampled permits against post office information file by finance number */
89
90 data merge1;
91   merge finance(in=a) sample(in=b);
92   by fin_num;
93   if b;
94 run;
```

NOTE: The data set WORK.MERGE1 has 995 observations and 20 variables.  
NOTE: The DATA statement used 0.82 seconds.

```
95
96 /* Merge sampled permits against customer information file by customer ID */
97
98 proc sort data=custinfo;
99   by cust_id;
100 run;
```

NOTE: The data set WORK.CUSTINFO has 52777 observations and 8 variables.  
NOTE: The PROCEDURE SORT used 5.76 seconds.

```
101
102 proc sort data=merge1;
103   by cust_id;
104 run;
```

NOTE: The data set WORK.MERGE1 has 995 observations and 20 variables.  
NOTE: The PROCEDURE SORT used 0.38 seconds.

```
105
106 data merge2;
107   merge custinfo(in=a) merge1(in=b);
108   by cust_id;
109   if b;
110 run;
```

NOTE: The data set WORK.MERGE2 has 995 observations and 27 variables.  
NOTE: The DATA statement used 1.37 seconds.

```
111
112 /* Calculate the average revenue per piece for the three categories by permit */
113
114 data sample;
115   set merge2;
116   if NAUTO_V = 0 then NAUTO_U = .;
117   else NAUTO_U = ROUND(NAUTO_R / NAUTO_V, 0.001);
118   if AUTO_V = 0 then AUTO_U = .;
119   else AUTO_U = ROUND(AUTO_R / AUTO_V, 0.001);
120   if CARD_V = 0 then CARD_U = .;
121   else CARD_U = ROUND(CARD_R / CARD_V, 0.001);
122 run;
```

NOTE: The data set WORK.SAMPLE has 995 observations and 30 variables.  
NOTE: The DATA statement used 0.55 seconds.

```
123
124 /* Print the sampled permits and associated information */
125
126 proc sort data=sample;
```

```
127 by descending strata sort_num;
128 run;
```

NOTE: The data set WORK.SAMPLE has 995 observations and 30 variables.  
NOTE: The PROCEDURE SORT used 0.22 seconds.

```
129
130 options nocenter nonumber nodate ls=255 ps=2500;
131
132 proc print data=sample split='*';
133 title 'Sample permits with customer and post office info';
134 var fin_num POcity POstate POzip permit indicia cust_id customer address
135 city state zip contact phone TOTAL_V NAUTO_U AUTO_U CARD_U NAUTO_R AUTO_R
136 CARD_R NAUTO_V AUTO_V CARD_V strata ran_num;
137 format bulk_vol NAUTO_V AUTO_V CARD_V NAUTO_R AUTO_R CARD_R comma12.
138 NAUTO_U AUTO_U CARD_U 6.3
WARNING: Variable BULK_VOL not found in data set WORK.SAMPLE.
139 ran_num 8.6;
140 label fin_num = 'FINANCE*NUMBER'
141 POcity = 'POST OFFICE*CITY'
142 POstate = 'POST*OFFICE*STATE'
143 POzip = 'POST*OFFICE*ZIP'
144 permit = 'PERMIT*NUMBER'
145 indicia = 'MAIL*TYPE'
146 cust_id = 'CUSTOMER*ID NUMBER'
147 customer = 'CUSTOMER NAME'
148 address = 'CUSTOMER ADDRESS'
149 city = 'CUSTOMER*CITY'
150 state = 'CUSTOMER*STATE'
151 zip = 'CUSTOMER*ZIP CODE'
152 contact = 'CUSTOMER CONTACT'
153 phone = 'CUSTOMER*PHONE'
154 TOTAL_V = 'TOTAL BULK*VOLUME'
155 NAUTO_U = 'CATEGORY 1*AVG PRICE'
156 AUTO_U = 'CATEGORY 2*AVG PRICE'
157 CARD_U = 'CATEGORY 3*AVG PRICE'
158 NAUTO_V = 'CATEGORY 1*VOLUME'
159 AUTO_V = 'CATEGORY 2*VOLUME'
160 CARD_V = 'CATEGORY 3*VOLUME'
161 NAUTO_R = 'CATEGORY 1*REVENUE'
162 AUTO_R = 'CATEGORY 2*REVENUE'
163 CARD_R = 'CATEGORY 3*REVENUE'
164 strata = 'STRATUM'
165 ran_num = 'RANDOM NUMBER*FOR SAMPLING';
166 run;
```

NOTE: The PROCEDURE PRINT used 1.04 seconds.

```
167
168 /* Output sampled permits and associated information to a text file */
169
170 data_NULL_;
171 set sample;
172 file out1;
173 put
174 @1 fin_num $6.
175 @8 POcity $15.
176 @24 POstate $2.
177 @27 POzip $5.
178 @34 permit $5.
179 @40 indicia $2.
180 @43 cust_id $10.
181 @54 customer $20.
182 @75 address $20.
183 @96 city $15.
184 @112 state $2.
185 @115 zip $5.
186 @121 contact $10.
187 @137 phone $10.
188 @148 TOTAL_V comma1.
189 @160 NAUTO_U 5.3
190 @166 AUTO_U 5.3
```

191 @172 CARD\_U 5.3  
192 @178 NAUTO\_R comma10.  
193 @189 AUTO\_R comma11.  
194 @201 CARD\_R comma9.  
195 @211 NAUTO\_V comma9.  
196 @221 AUTO\_V comma10.  
197 @232 CARD\_V comma8.  
198 @241 strata 1.  
199 @243 ran\_num 8.6;  
200 run;

NOTE: The file OUT1 is:  
FILENAME=D:\FCMDROP\FINAL.TXT,  
RECFM=V,LRECL=256

NOTE: 995 records were written to the file OUT1.

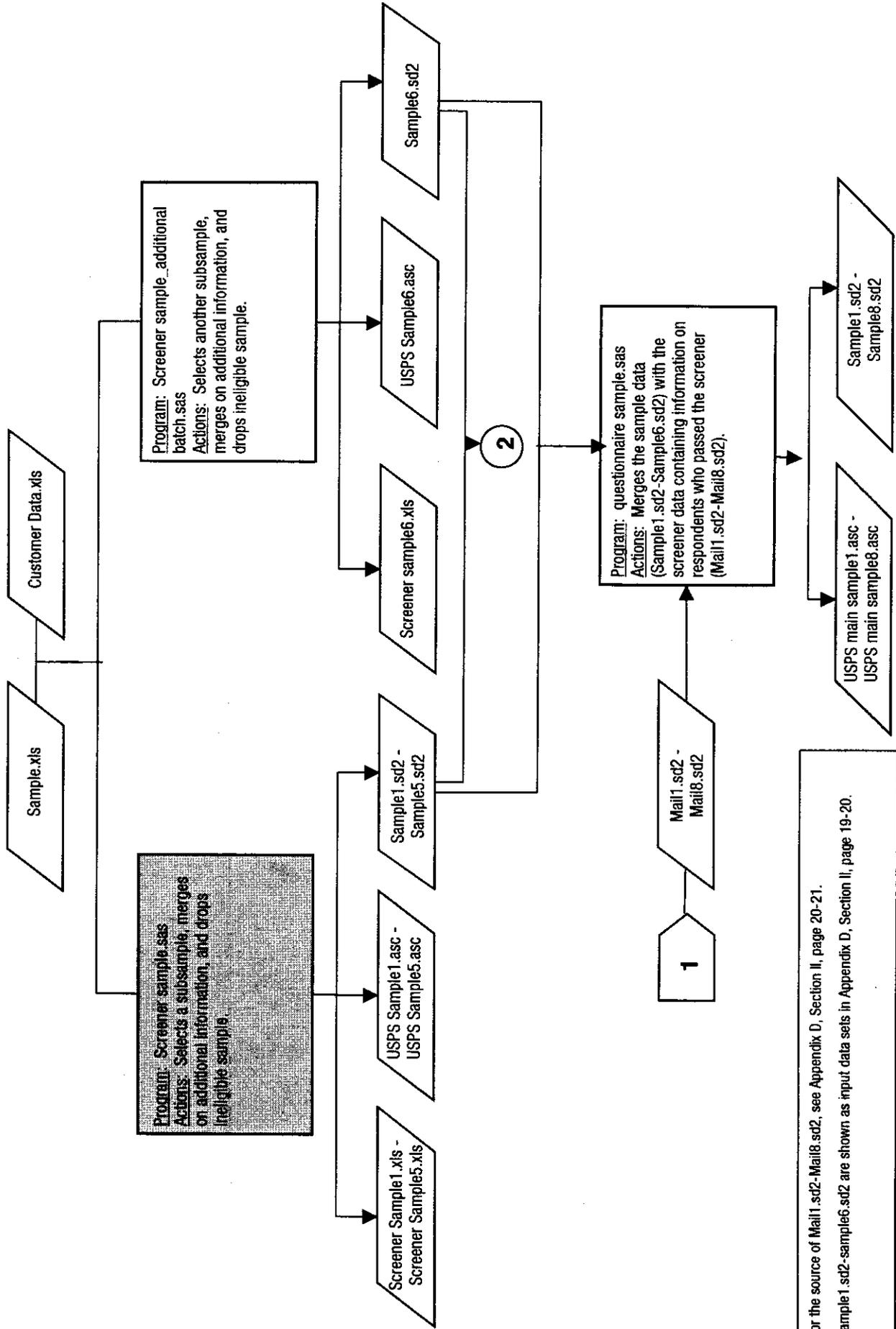
The minimum record length was 250.

The maximum record length was 250.

NOTE: The DATA statement used 0.55 seconds.

## **Section IV: Documentation of Screener Sample.sas**

# Flowchart of Sample Selection Programs



1 For the source of Mail1.sd2-Mail8.sd2, see Appendix D, Section II, page 20-21.  
 2 Sample1.sd2-sample6.sd2 are shown as input data sets in Appendix D, Section II, page 19-20.

## FIRST-CLASS MAIL SAMPLE SELECTION PROGRAMS

**STUDY:** USPS Destination-Entry Discount Study

**PROGRAM:** screener sample.sas

### I. Requirements of Computer Analysis Relied Upon

#### A. General description of the program:

1. Objectives of the program: The objectives of this program are to select a random subsample from each stratum from final.txt and merge on additional information from telephone calls to the issuing post offices.
2. Processing tasks performed: This program selects a random subsample within each stratum from final.txt. Permits owned by the Postal Service and USPS competitor companies are dropped. Additional information about the permits from calls to issuing post offices is merged with this subset of records. The subsample is output in batches because the additional information was provided on a rolling basis.
3. Methods and procedures employed: See attached program listing.
4. A listing of the input and output data: See attached pages describing names and sizes of input and output data.
5. A listing of source codes: See attached program and documentation.

#### B. Input data:

1. Designation of all sources of data: Sample.xls is an Excel spreadsheet created from final.txt, output from sample3.sas; customer data.xls is an Excel spreadsheet containing information gained from calls to issuing post offices. See Description of Variables section below for details on these new variables.
2. Explanation of any modifications of the data made for use in the program: The final.txt data set, produced by sample3.sas, was loaded into an Excel spreadsheet where five new variables were added to create the input for screener sample.sas, sample.xls. See Description of Variables section below for details on these new variables.

**C. Definitions of all input and output variables or sets of variables:** See attached description of input and output data.

**D. A description of input and output data file organization:** See attached description of input and output data.

**E. A machine-readable copy of all databases:** The databases used in this program contain confidential data and are not provided.

**F. For all source codes, documentation sufficiently comprehensive and detailed to satisfy generally accepted software documentation standards appropriate to the type of program and to its intended use in the proceedings:** See attached program and documentation.

**G. The source program in machine-readable form:** Both this program and the databases it uses are not provided due to confidentiality.

Name of Program: screener sample.sas

**H. All pertinent operating system and programming language manuals:** SAS Language: Reference, SAS Procedures Guide

**I. If requested program is interactive, a representative sample program run, together with any explanation necessary to illustrate the response sequence:** N/A

**J. "Canned" Statistical Packages:** SAS v6.12 for Windows

**K. Special requirements for computer simulations models offered if evidence or relied upon as support for other evidence:** N/A

**Program:** screener sample.sas

**Inputs:**

- (1) sample.xls  
Description: Sample.xls is an Excel file containing all variables and sampled permits from final.txt, output from sample3.sas.

Number of observations:	995
Number of variables:	31

- (2) customer data.xls  
Description: Customer data.xls is an Excel file containing information associated with each sampled permit obtained through calls to issuing post offices. Three of the permits were not able to be verified.

Number of observations:	992
Number of variables:	16

**Outputs:**

- (1) USPS sample1.asc  
Description: USPS sample1.asc is an ASCII file containing first batch of sampled permits for use in screening interviews.

Number of observations:	118
Number of variables:	21

- (2) USPS sample2.asc  
Description: USPS sample2.asc is an ASCII file containing second batch of sampled permits for use in screening interviews.

Number of observations:	78
Number of variables:	21

- (3) USPS sample3.asc  
Description: USPS sample3.asc is an ASCII file containing third batch of sampled permits for use in screening interviews.

Number of observations:	168
Number of variables:	21

- (4) USPS sample4.asc  
Description: USPS sample4.asc is an ASCII file containing fourth batch of sampled permits for use in screening interviews.

Number of observations:	29
Number of variables:	21

- (5) USPS sample5.asc  
Description: USPS sample5.asc is an ASCII file containing fifth batch of sampled permits for use in screening interviews.

Number of observations: 14  
Number of variables: 21

- (6) screener sample1.xls  
Description: Screener sample1.xls is an Excel file containing variables and sampled permits from USPS sample1.asc.

Number of observations: 118  
Number of variables: 43

- (7) screener sample2.xls  
Description: Sample2.xls is an Excel file containing variables and sampled permits from USPS sample2.asc.

Number of observations: 78  
Number of variables: 43

- (8) screener sample3.xls  
Description: Sample3.xls is an Excel file containing variables and sampled permits from USPS sample3.asc.

Number of observations: 168  
Number of variables: 43

- (9) screener sample4.xls  
Description: Sample4.xls is an Excel file containing variables and sampled permits from USPS sample4.asc.

Number of observations: 29  
Number of variables: 43

- (10) screener sample5.xls  
Description: Screener sample5.xls is an Excel file containing variables and sampled permits from USPS sample5.asc.

Number of observations: 14  
Number of variables: 43

- (11) sample1.sd2  
Description: Sample1.sd2 is a SAS data set containing variables and sampled permits from USPS sample1.asc.

Number of observations: 118  
Number of variables: 43

- (12) sample2.sd2

Description: Sample2.sd2 is a SAS data set containing variables and sampled permits from USPS sample2.asc.

Number of observations: 78  
Number of variables: 43

(13) sample3.sd2

Description: Sample3.sd2 is a SAS data set containing variables and sampled permits from USPS sample3.asc.

Number of observations: 168  
Number of variables: 43

(14) sample4.sd2

Description: Sample4.sd2 is a SAS data set containing variables and sampled permits from USPS sample4.asc.

Number of observations: 29  
Number of variables: 43

(15) sample5.sd2

Description: Sample5.sd2 is a SAS data set containing variables and sampled permits from USPS sample5.asc.

Number of observations: 14  
Number of variables: 43

#### Description of new variables in sample.xls:

OBS	Observation number from final.txt file
INTEST	Indicates if permit was also selected for pretest sample
COMP	Indicates if permit is owned by USPS or a USPS competitor company
MULTIPLE	Indicates if company at same address owns more than one permit in sample
CALLED	Indicates if company at same address was called during the pretest

#### Description of variables in customer data.xls:

FIN_NUM	Postal finance number
POCITY	Post office city
POSTATE	Post office state
POZIP	Post office ZIP Code (5-digit)
PERM NUM	Permit number
MAILTYPE	Permit indicia type (Metered, Permit Imprint, Precanceled Stamp)
CUSTNAME	Customer name
CUSTADD	Customer street address
CUSTCITY	Customer city
CUSSTATE	Customer state
CUSTZIP	Customer ZIP Code
NEWNAME	New company name gained from issuing post office calls
NEWCONT	New contact name gained from issuing post office calls
NEWPHONE	New contact telephone number gained from issuing post office calls
PERMDATE	Date permit was opened

NOTES Notes from calls to issuing post offices

**Description of new variables in: screener sample1.xls - screener sample5.xls  
sample1.sd2 - sample5.sd2:**

PWCID	Identification number for each sampled record
PERMTYPE	Indicia type for each sampled permit
COUNT1	Counter variables used to control selection of subsample in stratum 1.
COUNT2	Counter variables used to control selection of subsample in stratum 2.
COUNT3	Counter variables used to control selection of subsample in stratum 3.
ORDER	Random number assigned to each sampled record
COUNT	Number of permits owned by company with same customer identification number
CUSIDNUM	Customer identification number

Note: usps sample1.asc - usps sample5.asc contain a subset of the same variables found in sample1.sd2 - sample5.sd2.

**Actions of the program:**

- Assign identification number to each sampled record.
- If average price per piece is missing, set average price per piece to zero.
- If issuing post office state is missing, use post office ZIP Code to determine post office state.
- Create variable containing indicia type.
- Sort records by strata and random number and select a subset of records.
- Exclude USPS-owned and USPS competitor-owned permits.
- Create variable to count number of permits owned by each customer.
- Access permit information from Excel file and merge with sampled records.
- Keep records that have been verified through calls to the issuing post offices and produce a batch of these cases for use by the CATI system.

NOTE: Copyright (c) 1989-1996 by SAS Institute Inc., Cary, NC, USA.

NOTE: SAS (r) Proprietary Software Release 6.12 TS045  
Licensed to PRICEWATERHOUSECOOPERS LLP, Site 0015509006.

```
1 /*****
2   Program Name:  screener sample.sas
3   Author:       Rachel Allen
4   Date:         6/18/99
5   Last Modified: 7/22/99 (to produce batch 5)
6   Reviewed By:  Kelly Thomas on 7/22/99
7   Purpose:      To produce a text file for use in CATI system
8   Inputs:       sample.xls (Excel file containing 995 records sampled
9                from USPS database; customer data.xls)
10  Outputs:      usps sample1.asc - usps sample5.asc; screener sample1.xls - screener sample5.xls; sample1.sd2 - sample5.sd2
11  *****/
12
13  options ls=150 nodate nocenter;
14
15  libname usps 's:\ogs\common\usps\main study\screener\screener sample';
```

NOTE: Libref USPS was successfully assigned as follows:

Engine: V612  
Physical Name: s:\ogs\common\usps\main study\screener\screener sample

```
16
17
18  /*****
19  access sampled records from Excel file
20  *****/
21  proc access dbms=xls;
22      create usps.samp.access;
```

NOTE: Existing descriptor will be replaced at next CREATE or RUN.

```
23  path='s:\ogs\common\usps\main study\screener\screener sample\sample.xls';
24  worksheet='Sheet1';
25  range='a4:ae999';
26  getnames=yes;
27  mixed=yes;
28  type 1=c 2=c 3=c 4=c 5=c 6=c 7=c 8=c 9=c 10=c 11=c 12=c 13=c 14=c 15=c 16=n 17=n 18=n 19=n 20=n
29      21=n 22=n 23=n 24=n 25=n 26=n 27=n 28=n 29=n 30=n 31=n;
30  format obs $3. finnum $6. pocity $22. postate $2. pozip $5. permnum $5. mailtype $2. cusidnum $10.
31      custname $50. custadd $25. custcity $20. cusstate $2. custzip $5. custcont $20. custphon $10. totbulk comma15.2
32      cat1avg 5.3 cat2avg 5.3 cat3avg 5.3 cat1vol comma15.2 cat2vol comma15.2 cat3vol comma15.2 cat1rev comma15.2
33      cat2rev comma15.2 cat3rev comma15.2 strata 1. randnum 8.7 intest 1. comp 1. multiple 1. called 1.;
34  list all;
```

Excel File: S:\OGS\COMMON\USPS\MAIN Version: 5

Function: CREATE Descriptors- access: SAMP view:

Item	Column Label	SAS Name	Format
1	OBS		\$3.
2	FINNUM		\$6.
3	POCITY		\$22.
4	POSTATE		\$2.
5	POZIP		\$5.
6	PERMNUM		\$5.
7	MAILTYPE		\$2.
8	CUSIDNUM		\$10.
9	CUSTNAME		\$50.
10	CUSTADD		\$25.
11	CUSTCITY		\$20.
12	CUSSTATE		\$2.
13	CUSTZIP		\$5.
14	CUSTCONT		\$20.
15	CUSTPHON		\$10.
16	TOTBULK		COMMA15.2
17	CAT1AVG		5.3
18	CAT2AVG		5.3
19	CAT3AVG		5.3
20	CAT1VOL		COMMA15.2
21	CAT2VOL		COMMA15.2
22	CAT3VOL		COMMA15.2
23	CAT1REV		COMMA15.2

```

24 CAT2REV          COMMA15.2
25 CAT3REV          COMMA15.2
26 STRATA           1.0
27 RANDNUM          8.7
28 INTEST           1.0
29 COMP             1.0
30 MULTIPLE         1.0
31 CALLED           1.0

```

```
35      create usps.samp2.view;
```

NOTE: The access descriptor USPS.SAMP was written.

```
36      select all;
37      list all;
```

Excel File: SAOGS\COMMON\USPS\MAIN Version: 5  
Function: CREATE Descriptors- access: SAMP view: SAMP2

Item	Column Label	SAS Name	Format
1	OBS	OBS	\$3. *SELECTED *
2	FINNUM	FINNUM	\$6. *SELECTED *
3	POCITY	POCITY	\$22. *SELECTED *
4	POSTATE	POSTATE	\$2. *SELECTED *
5	POZIP	POZIP	\$5. *SELECTED *
6	PERMNUM	PERMNUM	\$5. *SELECTED *
7	MAILTYPE	MAILTYPE	\$2. *SELECTED *
8	CUSIDNUM	CUSIDNUM	\$10. *SELECTED *
9	CUSTNAME	CUSTNAME	\$50. *SELECTED *
10	CUSTADD	CUSTADD	\$25. *SELECTED *
11	CUSTCITY	CUSTCITY	\$20. *SELECTED *
12	CUSSTATE	CUSSTATE	\$2. *SELECTED *
13	CUSTZIP	CUSTZIP	\$5. *SELECTED *
14	CUSTCONT	CUSTCONT	\$20. *SELECTED *
15	CUSTPHON	CUSTPHON	\$10. *SELECTED *
16	TOTBULK	TOTBULK	COMMA15.2 *SELECTED *
17	CAT1AVG	CAT1AVG	5.3 *SELECTED *
18	CAT2AVG	CAT2AVG	5.3 *SELECTED *
19	CAT3AVG	CAT3AVG	5.3 *SELECTED *
20	CAT1VOL	CAT1VOL	COMMA15.2 *SELECTED *
21	CAT2VOL	CAT2VOL	COMMA15.2 *SELECTED *
22	CAT3VOL	CAT3VOL	COMMA15.2 *SELECTED *
23	CAT1REV	CAT1REV	COMMA15.2 *SELECTED *
24	CAT2REV	CAT2REV	COMMA15.2 *SELECTED *
25	CAT3REV	CAT3REV	COMMA15.2 *SELECTED *
26	STRATA	STRATA	1.0 *SELECTED *
27	RANDNUM	RANDNUM	8.7 *SELECTED *
28	INTEST	INTEST	1.0 *SELECTED *
29	COMP	COMP	1.0 *SELECTED *
30	MULTIPLE	MULTIPLE	1.0 *SELECTED *
31	CALLED	CALLED	1.0 *SELECTED *

```
38      run;
```

NOTE: The view descriptor USPS.SAMP2 was written.  
NOTE: The PROCEDURE ACCESS used 0.77 seconds.

```
39
40      proc access viewdesc=usps.samp2 out=samp;
```

NOTE: The data set WORK.SAMP has 995 observations and 31 variables.  
NOTE: The PROCEDURE ACCESS used 19.55 seconds.

```
41      run;
42
43
44
45      data samp;
46          set samp;
47
48      /******
49      use observation number as identification code (pwcid)
50      this will allow you to go back to the original sample file and
51          find the selected observation if needed
52      add 0's to front of pwcid so that it is 5 digits in length for use
```

```

53     in CATI center for look-ups
54     *****/
55     length pwcid $5.;
56     pwcid=obs;
57
58     if length(pwcid)=3 then pwcid='00'||pwcid;
59     else if length(pwcid)=2 then pwcid='000'||pwcid;
60     else if length(pwcid)=1 then pwcid='0000'||pwcid;
61
62     /*****
63     if average price per piece value is missing then average price is 0
64     *****/
65
66     if cat1avg=. then cat1avg=0;
67     if cat2avg=. then cat2avg=0;
68     if cat3avg=. then cat3avg=0;
69
70     /*****
71     if post office state is missing then use post office zip code to
72     determine state
73     *****/
74
75     if postate=" " then postate=zipstate(pozzip);
76
77     /*****
78     create permtype variable based on value of mailtype variable in sample
79     mailtype contains initials (i.e., PI) and permtype will contain
80     full word (i.e., permit imprinted mail)
81     *****/
82
83     if mailtype="PC" then permtype="Precanceled Stamp Mail";
84     else if mailtype="PI" then permtype="Permit Imprinted Mail";
85     else if mailtype="MT" then permtype="Metered Mail";
86 run;

```

NOTE: The data set WORK.SAMP has 995 observations and 33 variables.  
NOTE: The DATA statement used 0.33 seconds.

```

87
88
89
90     /*****
91     sort sampled records by strata and then by random number
92     note: random number was assigned during initial sample selection
93     *****/
94     proc sort data=samp;
95         by strata randnum;
96 run;

```

NOTE: The data set WORK.SAMP has 995 observations and 33 variables.  
NOTE: The PROCEDURE SORT used 0.22 seconds.

```

97
98
99
100    /*****
101    select sample for CATI study using following strata allocation:
102        small volume strata = 85 records (where strata=1)
103        medium volume strata = 131 records (where strata=2)
104        large volume strata = 160 records (where strata=3)
105        certainty strata = all (41) records (where strata=4)
106    *****/
107    data selected;
108    set samp;
109    if strata=1 then do;
110        count1+1;
111        if count1 <= 85;
112    end;
113    else if strata=2 then do;
114        count2+1;
115        if count2 <= 131;
116    end;
117    else if strata=3 then do;
118        count3+1;

```

```

119         if count3 <= 160;
120             end;
121         else if strata=4;

122 run;

```

NOTE: The data set WORK.SELECTED has 417 observations and 36 variables.  
NOTE: The DATA statement used 0.11 seconds.

```

123
124
125 data selected;
126     set selected;
127
128 /******
129 assign a new random number to each record
130 *****/
131
132     order=ranuni(400);
133
134 /******
135 exclude the following USPS-owned permits:
136 permnum mailtype custidnum custname
137 G0010 PI USPS USPS
138 G0010 PI USPS USPS
139 G0010 PI USPS USPS
140
141 exclude all USPS-competitor-owned permits
142 *****/
143
144     if comp=.;
145
146 /******
147 exclude the following permits that were selected in the pretest sample:
148 permnum mailtype custidnum custname
149 █████ MT █████ █████
150 █████ PI █████ █████
151 █████ MT █████ █████
152 █████ PI █████ █████
153 █████ MT █████ █████
154 █████ MT █████ █████
155 █████ MT █████ █████
156 *****/
157
158     if intest ne 1;
159 run;

```

NOTE: The data set WORK.SELECTED has 407 observations and 37 variables.  
NOTE: The DATA statement used 0.16 seconds.

```

160
161
162 /******
163 create count variable to indicate the number of
164 permits selected for a given USPS customer
165 *****/
166 proc freq data=selected;
167     tables cusidnum/out=counts (keep=cusidnum count);
168 run;

```

NOTE: The data set WORK.COUNTS has 387 observations and 2 variables.  
NOTE: The PROCEDURE FREQ printed pages 1-6.  
NOTE: The PROCEDURE FREQ used 0.27 seconds.

```

169
170
171
172 /******
173 merge count variable on to selected sample dataset
174 *****/
175 proc sort data=selected;
176     by cusidnum;
177 run;

```

NOTE: The data set WORK.SELECTED has 407 observations and 37 variables.  
NOTE: The PROCEDURE SORT used 0.11 seconds.

```
178
179 proc sort data=counts;
180     by cusidnum;
181 run;
```

NOTE: The data set WORK.COUNTS has 387 observations and 2 variables.  
NOTE: The PROCEDURE SORT used 0.05 seconds.

```
182
183 data selected;
184     merge selected counts;
185     by cusidnum;
186 run;
```

NOTE: The data set WORK.SELECTED has 407 observations and 38 variables.  
NOTE: The DATA statement used 0.11 seconds.

```
187
188
189
190 /*****
191 calls were placed to permit issuing post offices to get permit issue dates and owner information
192 after each day of verification calls, a spreadsheet was updated with the latest information
193 this SAS program was run each time the spreadsheet was revised to create a new batch of sample
194 *****/
195
196
197 /*****
198 access verified customer data and permit dates
199 *****/
200
201 proc access dbms=xls;
202     create usps.newdata.access;
```

NOTE: Existing descriptor will be replaced at next CREATE or RUN.

```
203     path='s:\ogs\common\usps\main study\screener\screener sample\customer data.xls';
204     worksheet='Sheet1';
205     range='a1:p993';
206     getnames=yes;
207     mixed=yes;
208     type 1=c 2=c 3=c 4=c 5=c 6=c 7=c 8=c 9=c 10=c 11=c 12=c 13=c 14=c 16=c;
209     format finnum $6. pocity $22. postate $2. pozip $5. permnum $5. mailtype $2. custname $50. custadd $25.
210     custcity $20. cusstate $2. custzip $5. newname $50. newcont $20. newphone $15. permdate mmddy8. notes $100.;
211     list all;
```

Excel File: S:\OGS\COMMON\USPS\MAIN Version: 5  
Function: CREATE Descriptors- access: NEWDATA view:

Item	Column Label	SAS Name Format
1	FINNUM	\$6.
2	POCITY	\$22.
3	POSTATE	\$2.
4	POZIP	\$5.
5	PERMNUM	\$5.
6	MAILTYPE	\$2.
7	CUSTNAME	\$50.
8	CUSTADD	\$25.
9	CUSTCITY	\$20.
10	CUSSTATE	\$2.
11	CUSTZIP	\$5.
12	NEWNAME	\$50.
13	NEWCONT	\$20.
14	NEWPHONE	\$15.
15	PERMDATE	MMDDYY8.
16	NOTES	\$100.

```
212     create usps.newdata2.view;
```

NOTE: The access descriptor USPS.NEWDATA was written.

```
213     select all;
214     list all;
```

Function: CREATE Descriptors- access: NEWDATA view: NEWDATA2

Item	Column Label	SAS Name	Format
1	FINNUM	FINNUM	\$6. * SELECTED *
2	POCITY	POCITY	\$22. * SELECTED *
3	POSTATE	POSTATE	\$2. * SELECTED *
4	POZIP	POZIP	\$5. * SELECTED *
5	PERMNUM	PERMNUM	\$5. * SELECTED *
6	MAILTYPE	MAILTYPE	\$2. * SELECTED *
7	CUSTNAME	CUSTNAME	\$50. * SELECTED *
8	CUSTADD	CUSTADD	\$25. * SELECTED *
9	CUSTCITY	CUSTCITY	\$20. * SELECTED *
10	CUSSTATE	CUSSTATE	\$2. * SELECTED *
11	CUSTZIP	CUSTZIP	\$5. * SELECTED *
12	NEWNAME	NEWNAME	\$50. * SELECTED *
13	NEWCONT	NEWCONT	\$20. * SELECTED *
14	NEWPHONE	NEWPHONE	\$15. * SELECTED *
15	PERMDATE	PERMDATE	MMDDYY8. * SELECTED *
16	NOTES	NOTES	\$100. * SELECTED *

215 run;

NOTE: The view descriptor USPS.NEWDATA2 was written.  
NOTE: The PROCEDURE ACCESS used 0.59 seconds.

216  
217 proc access viewdesc=usps.newdata2 out=newdata;  
NOTE: The data set WORK.NEWDATA has 992 observations and 16 variables.  
NOTE: The PROCEDURE ACCESS used 18.17 seconds.

218 run;  
219  
220  
221 /\*\*\*\*\*  
222 merge verified permit information (customer data.xls) onto selected sample  
223 \*\*\*\*\*/  
224  
225 proc sort data=selected;  
226 by finnum permnum mailtype custname custadd custcity custzip;  
227 run;

NOTE: The data set WORK.SELECTED has 407 observations and 38 variables.  
NOTE: The PROCEDURE SORT used 0.17 seconds.

228  
229 proc sort data=newdata;  
230 by finnum permnum mailtype custname custadd custcity custzip;  
231 run;

NOTE: The data set WORK.NEWDATA has 992 observations and 16 variables.  
NOTE: The PROCEDURE SORT used 0.11 seconds.

232  
233 data selected;  
234 merge selected (in=a) newdata (in=b);  
235 by finnum permnum mailtype custname custadd custcity custzip;  
236  
237  
238 /\*\*\*\*\*  
239 compress dashes (-) out of phone numbers  
240 if phone number is missing, put 2029999999  
241 \*\*\*\*\*/  
242  
243 newphone=compress(newphone, "-");  
244 if newphone=" " then newphone="2029999999";  
245  
246  
247 /\*\*\*\*\*  
248 if new owner informaton has been gathered, replace old info with new info from  
249 customer data.xls

```

250 because permit verification was done on a rolling basis, keep only permits
251 verified on the date that this program is run
252 *****/
253 if newname ne " " then custname=newname;
254 if newcont ne " " then custcont=newcont;
255 if newphone ne " " then custphon=newphone;
256 if a and b;
257 run;

```

NOTE: The data set WORK.SELECTED has 407 observations and 43 variables.  
NOTE: The DATA statement used 0.27 seconds.

```

258
259
260
261 *****/
262 keep only records from the latest batch of post office verification calls
263 *****/
264 proc sort data=usps.sample1;
265 by pwcid;
266 run;

```

NOTE: Input data set is already sorted, no sorting done.  
NOTE: The PROCEDURE SORT used 0.05 seconds.

```

267
268 proc sort data=usps.sample2;
269 by pwcid;
270 run;

```

NOTE: Input data set is already sorted, no sorting done.  
NOTE: The PROCEDURE SORT used 0.05 seconds.

```

271
272 proc sort data=usps.sample3;
273 by pwcid;
274 run;

```

NOTE: Input data set is already sorted, no sorting done.  
NOTE: The PROCEDURE SORT used 0.11 seconds.

```

275
276 proc sort data=usps.sample4;
277 by pwcid;
278 run;

```

NOTE: Input data set is already sorted, no sorting done.  
NOTE: The PROCEDURE SORT used 0.16 seconds.

```

279
280 proc sort data=selected;
281 by pwcid;
282 run;

```

NOTE: The data set WORK.SELECTED has 407 observations and 43 variables.  
NOTE: The PROCEDURE SORT used 0.17 seconds.

```

283
284 data selected;
285 merge usps.sample1 (in=a) usps.sample2 (in=b) usps.sample3 (in=c) usps.sample4 (in=d) selected (in=e);
286 by pwcid;
287 if not a and not b and not c and not d;
288 run;

```

NOTE: The data set WORK.SELECTED has 14 observations and 43 variables.  
NOTE: The DATA statement used 0.7 seconds.

```

289
290
291 *****/
292 randomly order the sampled units so that initial survey calls are not all made
293 to records in one stratum

```

```

294 *****/
295 proc sort data=selected;
296     by order;
297 run;

```

NOTE: The data set WORK.SELECTED has 14 observations and 43 variables.  
NOTE: The PROCEDURE SORT used 0.0 seconds.

```

298
299
300
301 /*****
302 create permanent data set containing selected records
303 each time a new batch is created, the usps.sample dataset will end in a
304     different number (e.g.,usps.sample1, usps.sample2)
305 *****/
306 data usps.sample5;
307     set selected;
308 run;

```

NOTE: The data set USPS.SAMPLE5 has 14 observations and 43 variables.  
NOTE: The DATA statement used 0.17 seconds.

```

309
310
311
312
313 /*****
314 output selected sample to Excel file
315 each time a new Excel file is created, the screener.sample.xls file will end in a
316     different number (e.g.,screener.sample1.xls, screener.sample2.xls)
317 *****/
318 proc dbload dbms=xls data=usps.sample5;
319     path="s:\ogs\common\usps\main study\screener\screener.sample5.xls";
320     putnames=yes;
321     list all;

```

--- Columns for: screener.sample5.xls ---

Num	Collabel	Type	SASname	Format
1	OBS	C	OBS	\$3.
2	FINNUM	C	FINNUM	\$6.
3	POCITY	C	POCITY	\$22.
4	POSTATE	C	POSTATE	\$2.
5	POZIP	C	POZIP	\$5.
6	PERMNUM	C	PERMNUM	\$5.
7	MAILTYPE	C	MAILTYPE	\$2.
8	CUSIDNUM	C	CUSIDNUM	\$10.
9	CUSTNAME	C	CUSTNAME	\$50.
10	CUSTADD	C	CUSTADD	\$25.
11	CUSTCITY	C	CUSTCITY	\$20.
12	CUSSTATE	C	CUSSTATE	\$2.
13	CUSTZIP	C	CUSTZIP	\$5.
14	CUSTCONT	C	CUSTCONT	\$20.
15	CUSTPHON	C	CUSTPHON	\$10.
16	TOTBULK	N	TOTBULK	COMMA15.2
17	CAT1AVG	N	CAT1AVG	5.3
18	CAT2AVG	N	CAT2AVG	5.3
19	CAT3AVG	N	CAT3AVG	5.3
20	CAT1VOL	N	CAT1VOL	COMMA15.2
21	CAT2VOL	N	CAT2VOL	COMMA15.2
22	CAT3VOL	N	CAT3VOL	COMMA15.2
23	CAT1REV	N	CAT1REV	COMMA15.2
24	CAT2REV	N	CAT2REV	COMMA15.2
25	CAT3REV	N	CAT3REV	COMMA15.2
26	STRATA	N	STRATA	1.
27	RANDNUM	N	RANDNUM	8.7
28	INTEST	N	INTEST	1.
29	COMP	N	COMP	1.
30	MULTIPLE	N	MULTIPLE	1.
31	CALLED	N	CALLED	1.
32	PWCID	C	PWCID	\$5.
33	PERMTYPE	C	PERMTYPE	\$22.
34	COUNTI	N	COUNTI	BEST15.

```

35 COUNT2          N COUNT2 BEST15.
36 COUNT3          N COUNT3 BEST15.
37 ORDER           N ORDER  BEST15.
38 COUNT           N COUNT  BEST15.
39 NEWNAME         C NEWNAME $50.
40 NEWCONT         C NEWCONT $20.
41 NEWPHONE        C NEWPHONE $15.
42 PERMDATE        N PERMDATE MMDDYY8.
43 NOTES           C NOTES  $100.

```

```

322 load;
323 run;

```

NOTE: Load completed. Examine statistics below.  
NOTE: Inserted (14) obs into screener sample5.xls.  
NOTE: Rejected (0) insert attempts see the log for details.

NOTE: The PROCEDURE DBLOAD used 1.04 seconds.

```

324
325
326 /*****
327 output CATI sample to fixed width ASCII file
328 each time a new ASCII file is created, the usps sample.asc file will end in a
329 different number (e.g.usps.sample1.asc, usps.sample2.asc)
330 *****/
331 data usps.sample5;
332 set usps.sample5;
333 file "s:\ogs\common\usps\main study\screener\screener sample\usps sample5.asc" lrecl=360;
334 put @1 custphon $10. /**permit-owner's phone number**/
335 @15 pwcid 5. /**identifying code**/
336 @20 custcont $20. /**company that owns permit**/
337 @45 custadd $25. /**company address**/
338 @75 custcity $20. /**company city**/
339 @100 cusstate $2. /**company state**/
340 @105 custzip $5. /**company zip**/
341 @115 custname $50. /**name of contact person at company**/
342 @170 permnum $5. /**FCM permit number**/
343 @180 permtype $29. /**type of mail: MT, PC or PI**/
344 @210 permdate mmddyy8. /**date that permit was issued**/
345 @220 totbulk comma15.2 /**total bulk volume for 1998**/
346 @240 cat1vol comma15.2 /**rate group 1 volume for 1998**/
347 @260 cat2vol comma15.2 /**rate group 2 volume for 1998**/
348 @280 cat3vol comma15.2 /**rate group 3 volume for 1998**/
349 @300 strata 1. /**1=small 2=medium 3=large 4=certainty**/
350 @305 pocity $22. /**city of issuing post office**/
351 @335 postate $2. /**state of issuing post office**/
352 @340 pozip $5. /**ZIP Code of issuing post office**/
353 @350 count 1. /**number of permits this company has in the sample**/
354 @355 called 1.; /**1=company was called in pretest about a different permit**/
355 run;

```

NOTE: The file "s:\ogs\common\usps\main study\screener\screener sample\usps sample5.asc" is:  
FILENAME=s:\ogs\common\usps\main study\screener\screener sample\usps sample5.asc,  
RECFM=V,LRECL=360

NOTE: 14 records were written to the file "s:\ogs\common\usps\main study\screener\screener sample\usps sample5.asc".  
The minimum record length was 355.

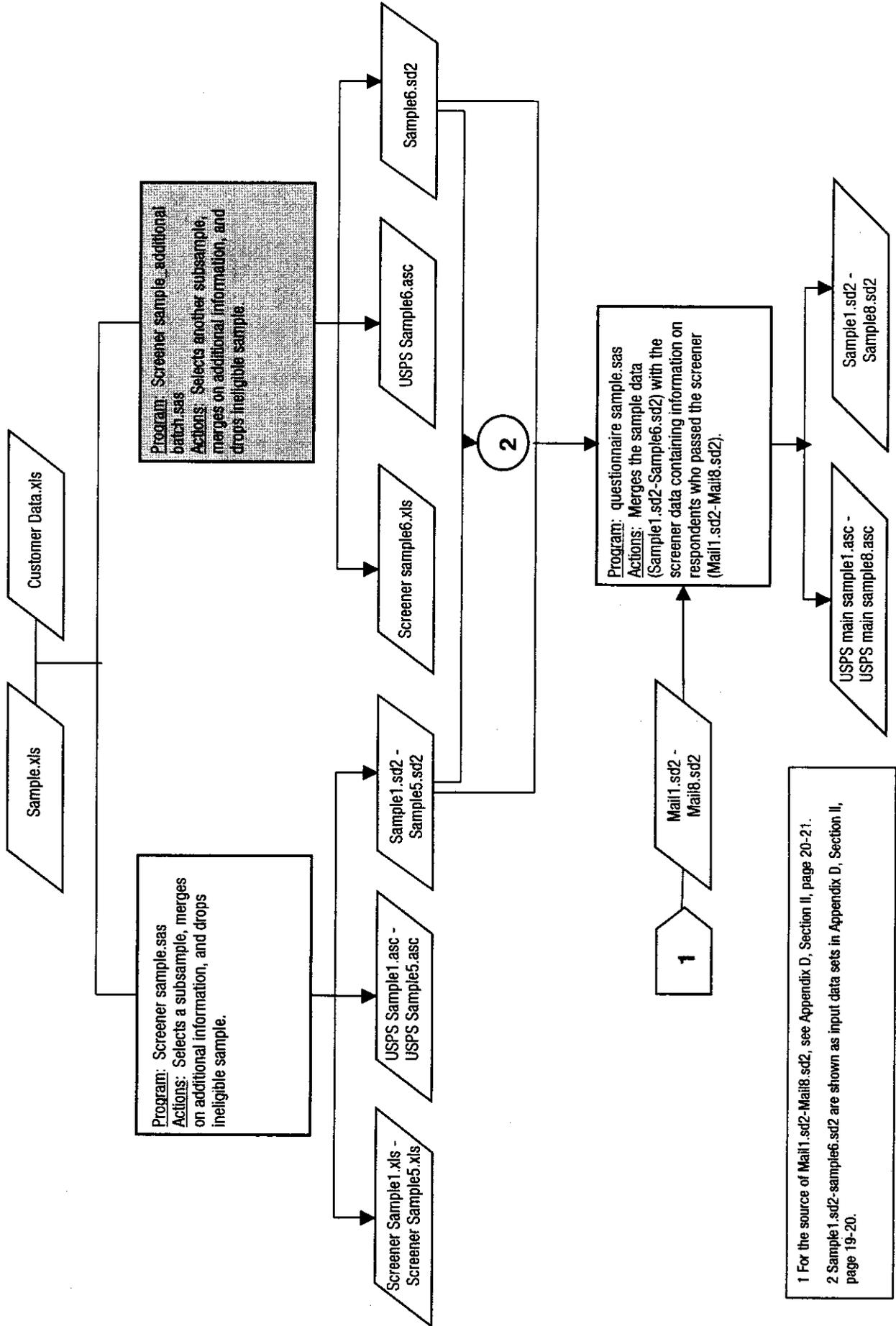
The maximum record length was 355.

NOTE: The data set USPS.SAMPLE5 has 14 observations and 43 variables.

NOTE: The DATA statement used 0.28 seconds.

## **Section V: Documentation of Screener Sample\_Additional Batch.sas**

# Flowchart of Sample Selection Programs



## FIRST-CLASS MAIL SAMPLE SELECTION PROGRAMS

**STUDY:** USPS Destination-Entry Discount Study  
**PROGRAM:** screener sample\_additional batch.sas

### I. Requirements of Computer Analysis Relied Upon

#### A. General description of the program:

1. Objectives of the program: The objective of this program is to select an additional subsample.
2. Processing tasks performed: This program selects an additional random subsample within strata 1-3 from final.txt. Permits owned by the Postal Service and USPS competitor companies are dropped. Additional information about the permits from calls to issuing post offices is merged with this subset of records.
3. Methods and procedures employed: See attached program listing.
4. A listing of the input and output data: See attached pages describing names and sizes of input and output data.
5. A listing of source codes: See attached program and documentation.

#### B. Input data:

1. Designation of all sources of data: Sample.xls is an Excel spreadsheet created from final.txt, output from sample3.sas; customer data.xls is an Excel spreadsheet containing information gained from calls to issuing post offices. See Description of Variables section below for details on these new variables.
2. Explanation of any modifications of the data made for use in the program: The final.txt data set, produced by sample3.sas, was loaded into an Excel spreadsheet where five new variables were added to create the input for screener sample.sas, sample.xls.

**C. Definitions of all input and output variables or sets of variables:** See attached description of input and output data.

**D. A description of input and output data file organization:** See attached description of input and output data.

**E. A machine-readable copy of all databases:** The databases used in this program contain confidential data and are not provided.

**F. For all source codes, documentation sufficiently comprehensive and detailed to satisfy generally accepted software documentation standards appropriate to the type of program and to its intended use in the proceedings:** See attached program and documentation.

**G. The source program in machine-readable form:** Both this program and the databases it uses are not provided due to confidentiality.

Name of Program: screener sample\_additional batch.sas

**H. All pertinent operating system and programming language manuals:** SAS Language: Reference, SAS Procedures Guide

- I. If requested program is interactive, a representative sample program run, together with any explanation necessary to illustrate the response sequence: N/A**
- J. "Canned" Statistical Packages: SAS v6.12 for Windows**
- K. Special requirements for computer simulations models offered if evidence or relied upon as support for other evidence: N/A**

**Program:** screener sample\_additional batch.sas

**Inputs:**

- (1) sample.xls  
Description: Sample.xls is an Excel file containing all variables and sampled permits from final.txt, output from sample3.sas.

Number of observations:	995
Number of variables:	31

- (2) customer data.xls  
Description: Customer data.xls is an Excel file containing variables associated with each sampled permit and variables obtained through calls to issuing post offices. Three of the permits were not able to be verified.

Number of observations:	992
Number of variables:	16

**Outputs:**

- (1) USPS sample6.asc  
Description: USPS sample6.asc is an ASCII file containing sixth batch of sampled permits for use in screening interviews.

Number of observations:	48
Number of variables:	21

- (2) screener sample6.xls  
Description: Sample6.xls is an Excel file containing variables and sampled permits from USPS sample6.asc.

Number of observations:	48
Number of variables:	43

- (3) sample6.sd2  
Description: Sample6.sd2 is a SAS data set containing variables and sampled permits from USPS sample6.asc.

Number of observations:	48
Number of variables:	43

**Actions of the program:**

- Assign identification number to each sampled record.
- If average price per piece is missing, set average price per piece to zero.
- If issuing post office state is missing, use post office ZIP Code to determine post office state.
- Create variable containing indicia type.
- Sort records by strata and random number and select a subset of records.
- Exclude USPS-owned and USPS competitor-owned permits.

- Create variable to count number of permits owned by each customer.
- Access permit information from Excel file and merge with sampled records.
- Keep records that have been verified through calls to the issuing post offices and produce a batch of these cases for use by the CATI system.

NOTE: Copyright (c) 1989-1996 by SAS Institute Inc., Cary, NC, USA.  
 NOTE: SAS (r) Proprietary Software Release 6.12 TS045  
 Licensed to PRICEWATERHOUSECOOPERS LLP, Site 0015509006.

```

1  /*****
2  Project:      USPS Destination-Entry FCM Study
3  Program Name: screener sample_additional batch.sas
4  Author:      Rachel Allen
5  Reviewed By: Kelly Thomas
6  Purpose:     To produce a text file for use in CATI system
7  Inputs:      sample.xls (Excel file containing 995 records sampled
8              from USPS database; customer data.xls)
9  Outputs:     usps sample6.asc (ASCII file for CATI)
10             screener sample6.xls (Excel file containing same records and variables as ASCII file)
11             sample6.sd2 (SAS file containing same records and variables as ASCII file)
12 *****/
13
14 options ls=150 nodate nocenter;
15
16 libname usps 's:\ogs\common\usps\main study\screener\screener sample';
  
```

NOTE: Libref USPS was successfully assigned as follows:  
 Engine: V612  
 Physical Name: s:\ogs\common\usps\main study\screener\screener sample

```

17
18
19 /*****
20 access sampled records from Excel file
21 *****/
22 proc access dbms=xls;
23   create usps.samp.access;
24   path='s:\ogs\common\usps\main study\screener\screener sample\sample.xls';
25   worksheet='Sheet1';
26   range='a4:ae999';
27   getnames=yes;
28   mixed=yes;
29   type 1=c 2=c 3=c 4=c 5=c 6=c 7=c 8=c 9=c 10=c 11=c 12=c 13=c 14=c 15=c 16=n 17=n 18=n 19=n 20=n
30       21=n 22=n 23=n 24=n 25=n 26=n 27=n 28=n 29=n 30=n 31=n;
31   format obs $3. finnum $6. pocity $22. postate $2. pozip $5. permnum $5. mailtype $2. cusidnum $10.
32         custname $50. custadd $25. custcity $20. cusstate $2. custzip $5. custcont $20. custphon $10. totbulk comma15.2
33         cat1avg 5.3 cat2avg 5.3 cat3avg 5.3 cat1vol comma15.2 cat2vol comma15.2 cat3vol comma15.2 cat1rev comma15.2
34         cat2rev comma15.2 cat3rev comma15.2 strata 1. randnum 8.7 intest 1. comp 1. multiple 1. called 1.;
35   list all;
  
```

Excel File: S:\OGS\COMMON\USPS\MAIN Version: 5  
 Function: CREATE Descriptors- access: SAMP view:

Item	Column Label	SAS Name	Format
1	OBS	\$3.	
2	FINNUM	\$6.	
3	POCITY	\$22.	
4	POSTATE	\$2.	
5	POZIP	\$5.	
6	PERMNUM	\$5.	
7	MAILTYPE	\$2.	
8	CUSIDNUM	\$10.	
9	CUSTNAME	\$50.	
10	CUSTADD	\$25.	
11	CUSTCITY	\$20.	
12	CUSSTATE	\$2.	
13	CUSTZIP	\$5.	
14	CUSTCONT	\$20.	
15	CUSTPHON	\$10.	
16	TOTBULK	COMMA15.2	
17	CAT1AVG	5.3	
18	CAT2AVG	5.3	
19	CAT3AVG	5.3	
20	CAT1VOL	COMMA15.2	
21	CAT2VOL	COMMA15.2	
22	CAT3VOL	COMMA15.2	
23	CATIREV	COMMA15.2	

```

24 CAT2REV          COMMA15.2
25 CAT3REV          COMMA15.2
26 STRATA           1.0
27 RANDNUM          8.7
28 INTEST           1.0
29 COMP             1.0
30 MULTIPLE         1.0
31 CALLED           1.0

```

```
36      create usps.samp2.view;
```

NOTE: The access descriptor USPS.SAMP was written.

```
37      select all;
38      list all;
```

Excel File: S:\OGS\COMMON\USPS\MAIN Version: 5  
Function: CREATE Descriptors- access: SAMP view: SAMP2

Item	Column Label	SAS Name	Format
1	OBS	OBS	\$3. * SELECTED *
2	FINNUM	FINNUM	\$6. * SELECTED *
3	POCITY	POCITY	\$22. * SELECTED *
4	POSTATE	POSTATE	\$2. * SELECTED *
5	POZIP	POZIP	\$5. * SELECTED *
6	PERMNUM	PERMNUM	\$5. * SELECTED *
7	MAILTYPE	MAILTYPE	\$2. * SELECTED *
8	CUSIDNUM	CUSIDNUM	\$10. * SELECTED *
9	CUSTNAME	CUSTNAME	\$50. * SELECTED *
10	CUSTADD	CUSTADD	\$25. * SELECTED *
11	CUSCITY	CUSCITY	\$20. * SELECTED *
12	CUSSTATE	CUSSTATE	\$2. * SELECTED *
13	CUSTZIP	CUSTZIP	\$5. * SELECTED *
14	CUSTCONT	CUSTCONT	\$20. * SELECTED *
15	CUSTPHON	CUSTPHON	\$10. * SELECTED *
16	TOTBULK	TOTBULK	COMMA15.2 * SELECTED *
17	CAT1AVG	CAT1AVG	5.3 * SELECTED *
18	CAT2AVG	CAT2AVG	5.3 * SELECTED *
19	CAT3AVG	CAT3AVG	5.3 * SELECTED *
20	CAT1VOL	CAT1VOL	COMMA15.2 * SELECTED *
21	CAT2VOL	CAT2VOL	COMMA15.2 * SELECTED *
22	CAT3VOL	CAT3VOL	COMMA15.2 * SELECTED *
23	CAT1REV	CAT1REV	COMMA15.2 * SELECTED *
24	CAT2REV	CAT2REV	COMMA15.2 * SELECTED *
25	CAT3REV	CAT3REV	COMMA15.2 * SELECTED *
26	STRATA	STRATA	1.0 * SELECTED *
27	RANDNUM	RANDNUM	8.7 * SELECTED *
28	INTEST	INTEST	1.0 * SELECTED *
29	COMP	COMP	1.0 * SELECTED *
30	MULTIPLE	MULTIPLE	1.0 * SELECTED *
31	CALLED	CALLED	1.0 * SELECTED *

```
39      run;
```

NOTE: The view descriptor USPS.SAMP2 was written.

NOTE: The PROCEDURE ACCESS used 8.17 seconds.

```
40
```

```
41      proc access viewdesc=usps.samp2 out=samp;
```

NOTE: The data set WORK.SAMP has 995 observations and 31 variables.

NOTE: The PROCEDURE ACCESS used 1 minute 23.48 seconds.

```
42      run;
```

```
43
```

```
44
```

```
45
```

```
46      data samp;
```

```
47      set samp;
```

```
48
```

```
49      /*****
```

```
50      use observation number as identification code (pwcid)
```

```
51      this will allow you to go back to the original sample file and
```

```
52      find the selected observation if needed
```

```
53      add 0's to front of pwcid so that it is 5 digits in length for use
```

```

54     in CATI center for look-ups
55     *****/
56     length pwcid $5.;
57     pwcid=obs;
58
59     if length(pwcid)=3 then pwcid='00'||pwcid;
60     else if length(pwcid)=2 then pwcid='000'||pwcid;
61     else if length(pwcid)=1 then pwcid='0000'||pwcid;
62
63     /*****
64     if average price per piece value is missing then average price is 0
65     *****/
66
67     if cat1avg=. then cat1avg=0;
68     if cat2avg=. then cat2avg=0;
69     if cat3avg=. then cat3avg=0;
70
71     /*****
72     if post office state is missing then use post office zip code to
73     determine state
74     *****/
75
76     if postate=" " then postate=zipstate(pozzip);
77
78     /*****
79     create permtype variable based on value of mailtype variable in sample
80     mailtype contains initials (i.e., PI) and permtype will contain
81     full word (i.e., permit imprinted mail)
82     *****/
83
84     if mailtype="PC" then permtype="Precanceled Stamp Mail";
85     else if mailtype="PI" then permtype="Permit Imprinted Mail";
86     else if mailtype="MT" then permtype="Metered Mail";
87 run;

```

NOTE: The data set WORK.SAMP has 995 observations and 33 variables.  
NOTE: The DATA statement used 0.38 seconds.

```

88
89
90
91     /*****
92     sort sampled records by strata and then by random number
93     note: random number was assigned during initial sample selection
94     *****/
95     proc sort data=samp;
96         by strata randnum;
97 run;

```

NOTE: The data set WORK.SAMP has 995 observations and 33 variables.  
NOTE: The PROCEDURE SORT used 0.27 seconds.

```

98
99
100
101     /*****
102     select additional sample for CATI screener using following strata allocation:
103         small bulk volume strata = 14 records (where strata=1)
104         medium bulk volume strata = 8 records (where strata=2)
105         large bulk volume strata = 26 records (where strata=3)
106         certainty strata = none
107     *****/
108     data selected;
109     set samp;
110     if strata=1 then do;
111         count1+1;
112         if count1 >=86 and count1 <100; ** note: records 1-85 were selected in the original screener sample **;
113     end;
114     else if strata=2 then do;
115         count2+1;
116         if count2 >=132 and count2 <140; ** note: records 1-131 were selected in the original screener sample **;
117     end;

```

```

118     else if strata=3 then do;
119         count3+1;
120         if count3 >=161 and count3 <187; ** note: records 1-160 were selected in the original screener sample **;
121         end;
122     if strata ne 4;
123 run;

```

NOTE: The data set WORK.SELECTED has 48 observations and 36 variables.  
NOTE: The DATA statement used 0.05 seconds.

```

124
125
126
127
128 data selected;
129     set selected;
130
131 /*****
132 assign a new random number to each record
133 *****/
134
135     order=ranuni(400);
136
137 /*****
138 exclude the following USPS competitor companies
139     or USPS-owned permits:
140 permnum mailtype custidnum custname
141 G0010 PI USPS USPS
142 G0010 PI USPS USPS
143 G0010 PI USPS USPS
144 *****/
145
146     if comp=.;
147
148 /*****
149 exclude the following permits that were selected in the pretest sample:
150 permnum mailtype custidnum custname
151 █████ MT █████ █████
152 █████ PI █████ █████
153 █████ MT █████ █████
154 █████ PI █████ █████
155 █████ MT █████ █████
156 █████ MT █████ █████
157 █████ MT █████ █████
158 *****/
159
160     if intest ne 1;
161 run;

```

NOTE: The data set WORK.SELECTED has 48 observations and 37 variables.  
NOTE: The DATA statement used 0.11 seconds.

```

162
163
164 /*****
165 create count variable to indicate the number of
166 permits selected for a given USPS customer
167 *****/
168 proc freq data=selected;
169     tables cusidnum/out=counts (keep=cusidnum count);
170 run;

```

NOTE: The data set WORK.COUNTS has 48 observations and 2 variables.  
NOTE: The PROCEDURE FREQ printed page 1.  
NOTE: The PROCEDURE FREQ used 0.38 seconds.

```

171
172
173
174 /*****
175 merge count variable on to selected sample dataset

```

```

176 *****/
177 proc sort data=selected;
178     by cusidnum;

179 run;

```

NOTE: The data set WORK.SELECTED has 48 observations and 37 variables.  
NOTE: The PROCEDURE SORT used 0.05 seconds.

```

180
181 proc sort data=counts;
182     by cusidnum;
183 run;

```

NOTE: The data set WORK.COUNTS has 48 observations and 2 variables.  
NOTE: The PROCEDURE SORT used 0.05 seconds.

```

184
185 data selected;
186     merge selected counts;
187     by cusidnum;
188 run;

```

NOTE: The data set WORK.SELECTED has 48 observations and 38 variables.  
NOTE: The DATA statement used 0.11 seconds.

```

189
190
191
192 /*****
193 calls were placed to permit issuing post offices to get permit issue dates and owner information
194 after each day of verification calls, a spreadsheet was updated with the latest information
195 this SAS program was run each time the spreadsheet was revised to create a new batch of sample
196 *****/
197
198
199 /*****
200 access verified customer data and permit dates
201 *****/
202
203 proc access dbms=xls;
204     create usps.newdata.access;

```

NOTE: Existing descriptor will be replaced at next CREATE or RUN.

```

205 path='s:\ogs\common\usps\main study\screener\screener sample\customer data.xls';
206 worksheet='Sheet1';
207 range='a1:p993';
208 getnames=yes;
209 mixed=yes;
210 type 1=c 2=c 3=c 4=c 5=c 6=c 7=c 8=c 9=c 10=c 11=c 12=c 13=c 14=c 16=c;
211 format finnum $6. pocity $22. postate $2. pozip $5. permnum $5. mailtype $2. custname $50. custadd $25.
212     custcity $20. cusstate $2. custzip $5. newname $50. newcont $20. newphone $15. permdate mmdyy8. notes $100.;
213 list all;

```

Excel File: S:\OGS\COMMON\USPS\MAIN Version: 5  
Function: CREATE Descriptors- access: NEWDATA view:

Item	Column Label	SAS Name Format
1	FINNUM	\$6.
2	POCITY	\$22.
3	POSTATE	\$2.
4	POZIP	\$5.
5	PERMNUM	\$5.
6	MAILTYPE	\$2.
7	CUSTNAME	\$50.
8	CUSTADD	\$25.
9	CUSTCITY	\$20.
10	CUSSTATE	\$2.
11	CUSTZIP	\$5.
12	NEWNAME	\$50.
13	NEWCONT	\$20.
14	NEWPHONE	\$15.

15 PERMDATE MMDDYY8.  
16 NOTES \$100.

214 create usps.newdata2.view;

NOTE: The access descriptor USPS.NEWDATA was written.

215 select all;  
216 list all;

Excel File: S:\OGS\COMMON\USPS\MAIN Version: 5  
Function: CREATE Descriptors- access: NEWDATA view: NEWDATA2

Item	Column Label	SAS Name	Format
1	FINNUM	FINNUM \$6.	* SELECTED *
2	POCITY	POCITY \$22.	* SELECTED *
3	POSTATE	POSTATE \$2.	* SELECTED *
4	POZIP	POZIP \$5.	* SELECTED *
5	PERMNUM	PERMNUM \$5.	* SELECTED *
6	MAILTYPE	MAILTYPE \$2.	* SELECTED *
7	CUSTNAME	CUSTNAME \$50.	* SELECTED *
8	CUSTADD	CUSTADD \$25.	* SELECTED *
9	CUSTCITY	CUSTCITY \$20.	* SELECTED *
10	CUSSTATE	CUSSTATE \$2.	* SELECTED *
11	CUSTZIP	CUSTZIP \$5.	* SELECTED *
12	NEWNAME	NEWNAME \$50.	* SELECTED *
13	NEWCONT	NEWCONT \$20.	* SELECTED *
14	NEWPHONE	NEWPHONE \$15.	* SELECTED *
15	PERMDATE	PERMDATE MMDDYY8.	* SELECTED *
16	NOTES	NOTES \$100.	* SELECTED *

217 run;

NOTE: The view descriptor USPS.NEWDATA2 was written.  
NOTE: The PROCEDURE ACCESS used 5.33 seconds.

218  
219 proc access viewdesc=usps.newdata2 out=newdata;  
NOTE: The data set WORK.NEWDATA has 992 observations and 16 variables.  
NOTE: The PROCEDURE ACCESS used 1 minute 16.73 seconds.

220 run;  
221  
222  
223 /\*\*\*\*\*  
224 merge verified permit information (customer data.xls) onto selected sample  
225 \*\*\*\*\*/  
226  
227 proc sort data=selected;  
228 by finnum permnum mailtype custname custadd custcity custzip;  
229 run;

NOTE: The data set WORK.SELECTED has 48 observations and 38 variables.  
NOTE: The PROCEDURE SORT used 0.05 seconds.

230  
231 proc sort data=newdata;  
232 by finnum permnum mailtype custname custadd custcity custzip;  
233 run;

NOTE: The data set WORK.NEWDATA has 992 observations and 16 variables.  
NOTE: The PROCEDURE SORT used 0.17 seconds.

234  
235 data selected;  
236 merge selected (in=a) newdata (in=b);  
237 by finnum permnum mailtype custname custadd custcity custzip;  
238  
239  
240 /\*\*\*\*\*  
241 compress dashes (-) out of phone numbers

```

242 if phone number is missing, put 2029999999
243 *****/
244
245     newphone=compress(newphone,"-");
246     if newphone=" " then newphone="2029999999";
247
248
249 /*****
250 if new owner informaton has been gathered, replace old info with new info from
251     customer data.xls
252 because permit verification was done on a rolling basis, keep only permits
253     verified on the date that this program is run
254 *****/
255     if newname ne " " then custname=newname;
256     if newcont ne " " then custcont=newcont;
257     if newphone ne " " then custphon=newphone;
258     if a and b;
259 run;

```

NOTE: The data set WORK.SELECTED has 48 observations and 43 variables.  
NOTE: The DATA statement used 0.33 seconds.

```

260
261
262
263 /*****
264 randomly order the sampled units so that initial survey calls are not all made
265 to records in one stratum
266 *****/
267 proc sort data=selected;
268     by order;
269 run;

```

NOTE: The data set WORK.SELECTED has 48 observations and 43 variables.  
NOTE: The PROCEDURE SORT used 0.05 seconds.

```

270
271
272
273 /*****
274 create permanent data set containing selected records
275 *****/
276 data usps.sample6;
277     set selected;
278 run;

```

NOTE: The data set USPS.SAMPLE6 has 48 observations and 43 variables.  
NOTE: The DATA statement used 0.28 seconds.

```

279
280
281
282
283 /*****
284 output selected sample to Excel file
285 *****/
286 proc dbload dbms=xls data=usps.sample6;
287     path="s:\logs\common\usps\main study\screener\screener sample\screener sample6.xls";
288     putnames=yes;
289 list all;

```

--- Columns for: screener sample6.xls ---

Num	Collabel	Type	SASname	Format
1	OBS	C	OBS	\$3.
2	FINNUM	C	FINNUM	\$6.
3	POCITY	C	POCITY	\$22.
4	POSTATE	C	POSTATE	\$2.
5	POZIP	C	POZIP	\$5.
6	PERMNUM	C	PERMNUM	\$5.
7	MAILTYPE	C	MAILTYPE	\$2.
8	CUSIDNUM	C	CUSIDNUM	\$10.

```

9  CUSTNAME          C CUSTNAME $50.
10 CUSTADD           C CUSTADD $25.
11 CUSTCITY          C CUSTCITY $20.
12 CUSSTATE          C CUSSTATE $2.
13 CUSTZIP           C CUSTZIP $5.
14 CUSTCONT          C CUSTCONT $20.
15 CUSTPHON          C CUSTPHON $10.
16 TOTBULK           N TOTBULK COMMA15.2
17 CAT1AVG           N CAT1AVG 5.3
18 CAT2AVG           N CAT2AVG 5.3
19 CAT3AVG           N CAT3AVG 5.3
20 CAT1VOL           N CAT1VOL COMMA15.2
21 CAT2VOL           N CAT2VOL COMMA15.2
22 CAT3VOL           N CAT3VOL COMMA15.2
23 CAT1REV           N CAT1REV COMMA15.2
24 CAT2REV           N CAT2REV COMMA15.2
25 CAT3REV           N CAT3REV COMMA15.2
26 STRATA            N STRATA 1.
27 RANDNUM           N RANDNUM 8.7
28 INTEST            N INTEST 1.
29 COMP              N COMP 1.
30 MULTIPLE          N MULTIPLE 1.
31 CALLED            N CALLED 1.
32 PWCID             C PWCID $5.
33 PERMTYPE          C PERMTYPE $22.
34 COUNT1            N COUNT1 BEST15.
35 COUNT2            N COUNT2 BEST15.
36 COUNT3            N COUNT3 BEST15.
37 ORDER             N ORDER BEST15.
38 COUNT             N COUNT BEST15.
39 NEWNAME           C NEWNAME $50.
40 NEWCONT           C NEWCONT $20.
41 NEWPHONE          C NEWPHONE $15.
42 PERMDATE          N PERMDATE MMDDYY8.
43 NOTES             C NOTES $100.

```

```

290 load;
291 run;

```

NOTE: Load completed. Examine statistics below.  
NOTE: Inserted (48) obs into screener sample6.xls.  
NOTE: Rejected (0) insert attempts see the log for details.

NOTE: The PROCEDURE DBLOAD used 9.99 seconds.

```

292
293
294 /******
295 output CATI sample to fixed width ASCII file
296 *****/
297 data usps.sample6;
298   set usps.sample6;
299   file "s:\logs\common\usps\main study\screener\screener sample\usps sample6.asc" irecl=360;
300   put @1 custphon $10.    /**permit-owner's phone number**/
301     @15 pwcid 5.         /**identifying code**/
302     @20 custcont $20.    /**company that owns permit**/
303     @45 custadd $25.     /**company address**/
304     @75 custcity $20.    /**company city**/
305     @100 cusstate $2.    /**company state**/
306     @105 custzip $5.     /**company zip**/
307     @115 custname $50.   /**name of contact person at company**/
308     @170 permnum $5.     /**FCM permit number**/
309     @180 permtype $29.   /**type of mail: MT, PC or PI**/
310     @210 permdate mmddyy8. /**date that permit was issued**/
311     @220 totbulk comma15.2 /**total bulk volume for 1998**/
312     @240 cat1vol comma15.2 /**rate group 1 volume for 1998**/
313     @260 cat2vol comma15.2 /**rate group 2 volume for 1998**/
314     @280 cat3vol comma15.2 /**rate group 3 volume for 1998**/
315     @300 strata 1.       /**1=small 2=medium 3=large 4=certainty**/
316     @305 pocity $22.     /**city of issuing post office**/
317     @335 postate $2.     /**state of issuing post office**/
318     @340 pozzip $5.      /**ZIP Code of issuing post office**/
319     @350 count 1.       /**number of permits this company has in the sample**/

```

320            @355 called 1.;     /\*\*1=company was called in pretest about a different permit\*\*/  
321     run;

NOTE: The file "s:\logs\common\usps\main study\screeener\screeener sample\usps sample6.asc" is:  
    FILENAME=s:\logs\common\usps\main study\screeener\screeener sample\usps sample6.asc.  
    RECFM=V,LRECL=360

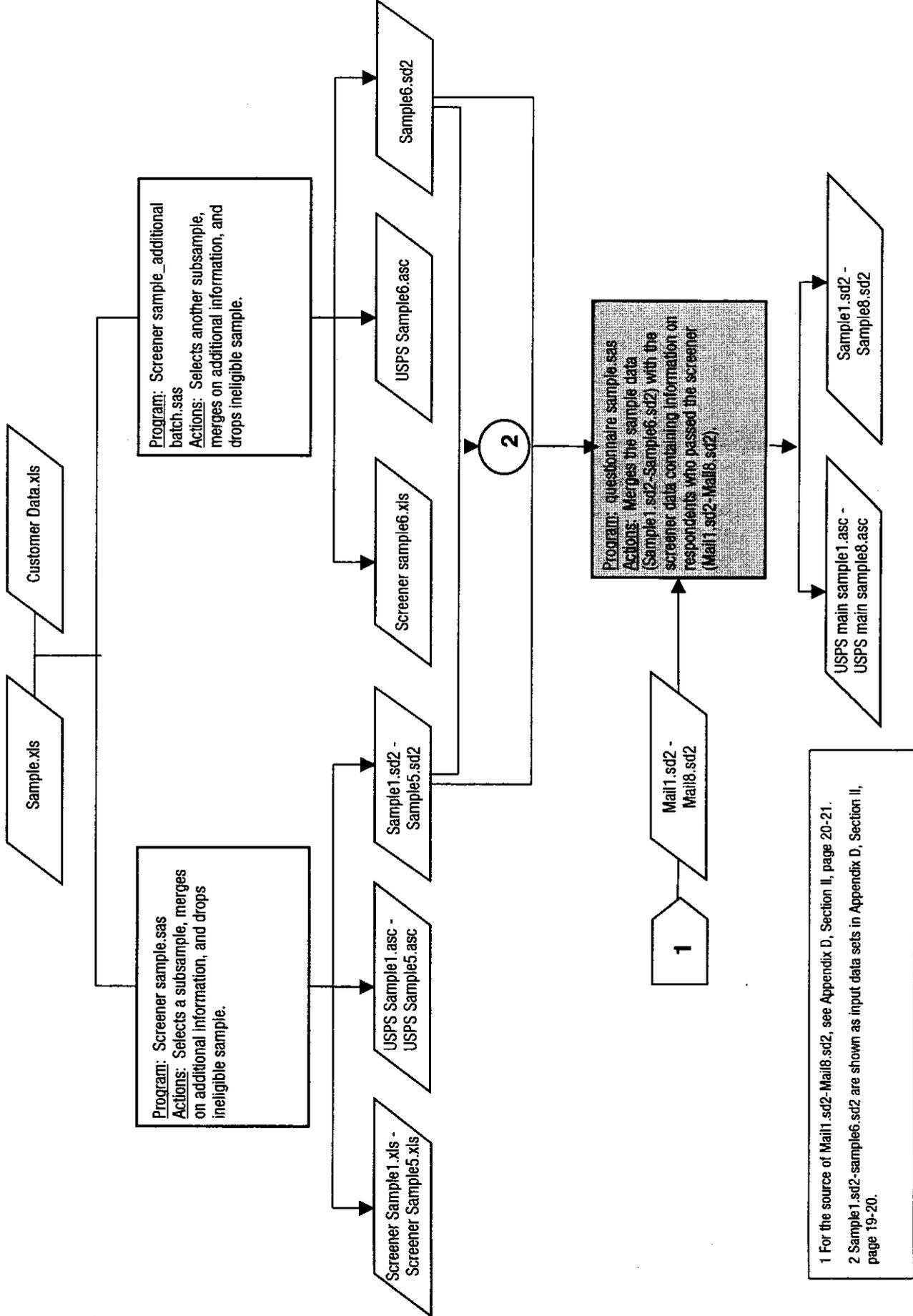
NOTE: 48 records were written to the file "s:\logs\common\usps\main study\screeener\screeener sample\usps sample6.asc".  
    The minimum record length was 355.  
    The maximum record length was 355.

NOTE: The data set USPS.SAMPLE6 has 48 observations and 43 variables.

NOTE: The DATA statement used 5.87 seconds.

## **Section VI: Documentation of Questionnaire Sample.sas**

# Flowchart of Sample Selection Programs



## FIRST-CLASS MAIL SAMPLE SELECTION PROGRAMS

**STUDY:** USPS Destination-Entry Discount Study

**PROGRAM:** questionnaire sample.sas

### I. Requirements of Computer Analysis Relied Upon

#### A. General description of the program:

1. Objectives of the program: The objective of this program is to merge the sample data with the data containing information on respondents who passed the screener to create a CATI file for the survey questionnaire.
2. Processing tasks performed: This program merges the sample data with the data containing information on respondents who passed the screener to create a CATI file for the survey questionnaire. This program was run eight separate times to create a separate batch for each of the eight mailings.
3. Methods and procedures employed: See attached program listing.
4. A listing of the input and output data: See attached pages describing names and sizes of input and output data.
5. A listing of source codes: See attached program and documentation.

#### B. Input data:

1. Designation of all sources of data: Sample1.sd2 - sample5.sd2 are output from screener sample.sas; sample6.sd2 is output from screener sample\_additional batch.sas; mail1.sd2 - mail8.sd2 are output from mailpak.sas. For documentation of the program, mailpak.sas, see Section II of Appendix D.
2. Explanation of any modifications of the data made for use in the program: None.

**C. Definitions of all input and output variables or sets of variables:** See attached description of input and output data.

**D. A description of input and output data file organization:** See attached description of input and output data.

**E. A machine-readable copy of all databases:** The databases used in this program contain confidential data and are not provided.

**F. For all source codes, documentation sufficiently comprehensive and detailed to satisfy generally accepted software documentation standards appropriate to the type of program and to its intended use in the proceedings:** See attached program and documentation.

**G. The source program in machine-readable form:** Both this program and the databases it uses are not provided due to confidentiality. Since the program alone does not contain identifying information, it can be provided upon request.

Name of Program: questionnaire sample.sas

- H. All pertinent operating system and programming language manuals: SAS Language: Reference, SAS Procedures Guide**
- I. If requested program is interactive, a representative sample program run, together with any explanation necessary to illustrate the response sequence: N/A**
- J. "Canned" Statistical Packages: SAS v6.12 for Windows**
- K. Special requirements for computer simulations models offered if evidence or relied upon as support for other evidence: N/A**

**Program:** questionnaire sample.sas

**Inputs:**

- (1) sample1.sd2  
Description: Sample1.sd2 is a SAS data set containing the first batch of screening interview sample; output from screener sample.sas.  

Number of observations:	118
Number of variables:	43
  
- (2) sample2.sd2  
Description: Sample2.sd2 is a SAS data set containing the second batch of screening interview sample; output from screener sample.sas.  

Number of observations:	78
Number of variables:	43
  
- (3) sample3.sd2  
Description: Sample3.sd2 is a SAS data set containing the third batch of screening interview sample; output from screener sample.sas.  

Number of observations:	168
Number of variables:	43
  
- (4) sample4.sd2  
Description: Sample4.sd2 is a SAS data set containing the fourth batch of screening interview sample; output from screener sample.sas.  

Number of observations:	29
Number of variables:	43
  
- (5) sample5.sd2  
Description: Sample5.sd2 is a SAS data set containing the fifth batch of screening interview sample; output from screener sample.sas.  

Number of observations:	14
Number of variables:	43
  
- (6) sample6.sd2  
Description: Sample6.sd2 is a SAS data set containing the sixth batch of screening interview sample; output from screener sample\_additional batch.sas.  

Number of observations:	48
Number of variables:	43
  
- (7) mail1.sd2  
Description: Mail1.sd2 is a SAS data set (produced by mailpak.sas) of survey cases comprising the first mailing.

Number of observations: 83  
Number of variables: 22

- (8) mail2.sd2  
Description: Mail2.sd2 is a SAS data set (produced by mailpak.sas) of survey cases comprising the second mailing.

Number of observations: 87  
Number of variables: 22

- (9) mail3.sd2  
Description: Mail3.sd2 is a SAS data set (produced by mailpak.sas) of survey cases comprising the third mailing.

Number of observations: 45  
Number of variables: 22

- (10) mail4.sd2  
Description: Mail4.sd2 is a SAS data set (produced by mailpak.sas) of survey cases comprising the fourth mailing.

Number of observations: 22  
Number of variables: 22

- (11) mail5.sd2  
Description: Mail5.sd2 is a SAS data set (produced by mailpak.sas) of survey cases comprising the fifth mailing.

Number of observations: 65  
Number of variables: 22

- (12) mail6.sd2  
Description: Mail6.sd2 is a SAS data set (produced by mailpak.sas) of survey cases comprising the sixth mailing.

Number of observations: 18  
Number of variables: 22

- (13) mail7.sd2  
Description: Mail7.sd2 is a SAS data set (produced by mailpak.sas) of survey cases comprising the seventh mailing.

Number of observations: 21  
Number of variables: 22

- (14) mail8.sd2

Description: Mail8.sd2 is a SAS data set (produced by mailpak.sas) of survey cases comprising the eighth mailing.

Number of observations:	4
Number of variables:	22

**Outputs:**

- (1) USPS main sample1.asc  
Description: USPS main sample1.asc is an ASCII file containing permits from first mailout for use in main survey interviews.  

Number of observations:	83
Number of variables:	21
  
- (2) USPS main sample2.asc  
Description: USPS main sample2.asc is an ASCII file containing permits from second mailout for use in main survey interviews.  

Number of observations:	87
Number of variables:	21
  
- (3) USPS main sample3.asc  
Description: USPS main sample3.asc is an ASCII file containing permits from third mailout for use in main survey interviews.  

Number of observations:	45
Number of variables:	21
  
- (4) USPS main sample4.asc  
Description: USPS main sample4.asc is an ASCII file containing permits from fourth mailout for use in main survey interviews.  

Number of observations:	22
Number of variables:	21
  
- (5) USPS main sample5.asc  
Description: USPS main sample5.asc is an ASCII file containing permits from fifth mailout for use in main survey interviews.  

Number of observations:	65
Number of variables:	21
  
- (6) USPS main sample6.asc  
Description: USPS main sample6.asc is an ASCII file containing permits from sixth mailout for use in main survey interviews.  

Number of observations:	18
-------------------------	----

Number of variables: 21

- (7) USPS main sample7.asc  
Description: USPS main sample7.asc is an ASCII file containing permits from seventh mailout for use in main survey interviews.

Number of observations: 21  
Number of variables: 21

- (8) USPS main sample8.asc  
Description: USPS main sample8.asc is an ASCII file containing permits from eighth mailout for use in main survey interviews.

Number of observations: 4  
Number of variables: 21

- (9) sample1.sd2  
Description: Sample1.sd2 is a SAS data set containing permits from first mailout.

Number of observations: 83  
Number of variables: 43

- (10) sample2.sd2  
Description: Sample2.sd2 SAS data set containing permits from second mailout.

Number of observations: 87  
Number of variables: 43

- (11) sample3.sd2  
Description: Sample3.sd2 is a SAS data set containing permits from third mailout.

Number of observations: 45  
Number of variables: 43

- (12) sample4.sd2  
Description: Sample4.sd2 is a SAS data set containing permits from fourth mailout.

Number of observations: 22  
Number of variables: 43

- (13) sample5.sd2  
Description: Sample5.sd2 is a SAS data set containing permits from fifth mailout.

Number of observations: 65  
Number of variables: 43

- (14) sample6.sd2  
Description: Sample6.sd2 is a SAS data set containing permits from sixth mailout.

Number of observations:	18
Number of variables:	43

- (15) sample7.sd2  
Description: Sample7.sd2 is a SAS data set containing permits from seventh mailout.

Number of observations:	21
Number of variables:	43

- (16) sample8.sd2  
Description: Sample8.sd2 is a SAS data set containing permits from eighth mailout.

Number of observations:	4
Number of variables:	43

**Actions of the program:**

- Merge the sample data with the data containing information on respondents who passed the screener.

NOTE: Copyright (c) 1989-1996 by SAS Institute Inc., Cary, NC, USA.  
NOTE: SAS (r) Proprietary Software Release 6.12 TS045  
Licensed to PRICEWATERHOUSECOOPERS LLP, Site 0015509006.

```
1  /*****
2  Project:      USPS Destination-Entry FCM Study
3  Program Name: questionnaire sample.sas
4  Author:       Rachel Allen
5  Reviewed By:  Kelly Thomas
6  Purpose:      To produce a text file for use in CATI system
7  Inputs:       sample1.sd2 sample2.sd2 sample3.sd2 sample4.sd2 sample5.sd2 sample6.sd2
8               (6 batches of sample used in screening questionnaire)
9               mail1.sd2 - mail8.sd2 (dataset containing permits in each mailout)
10 Outputs:      usps main sample1.txt - usps main sample8.txt (text files for CATI)
11               sample1.sd2 - sample8.sd2
12 *****/
13
14 options ls=150 nodate nocenter;
15
16 libname screen 's:\logs\common\usps\main study\screener\screener sample';
```

NOTE: Libref SCREEN was successfully assigned as follows:

```
Engine: V612
Physical Name: s:\logs\common\usps\main study\screener\screener sample
```

```
17 libname main 's:\logs\common\usps\main study\questionnaire\questionnaire sample';
```

NOTE: Libref MAIN was successfully assigned as follows:

```
Engine: V612
Physical Name: s:\logs\common\usps\main study\questionnaire\questionnaire sample
```

```
18 libname mailed 's:\logs\common\usps\main study\screener\screener data';
```

NOTE: Libref MAILED was successfully assigned as follows:

```
Engine: V612
Physical Name: s:\logs\common\usps\main study\screener\screener data
```

```
19
20
21 %let i=8; /** each time this program is run, the i value should be incremented by 1 **/
22
23
24
25 /*****
26 access sampled records from screener sample batches
27 sort by pwcid
28 drop customer name and contact information from this dataset because the
29 correct information will be picked up in the mailed dataset
30 *****/
31 data samples (drop=custphon custcont custname custadd custcity cusstate custzip);
32 set screen.sample1 screen.sample2 screen.sample3 screen.sample4 screen.sample5 screen.sample6;
33 run;
```

NOTE: The data set WORK.SAMPLES has 455 observations and 36 variables.

NOTE: The DATA statement used 4.17 seconds.

```
34
35
36 proc sort data=samples;
37 by pwcid;
38 run;
```

NOTE: The data set WORK.SAMPLES has 455 observations and 36 variables.

NOTE: The PROCEDURE SORT used 0.16 seconds.

```
39
40
41
42 /*****
43 access records from last mailout
```

```

44  sort by pwcid
45  keep only pwcid and contact information
46  *****/
47  data mailed (keep=pwcid address city company contact state zip phone1);
48  set mailed.mail&i;
49  run;

```

NOTE: The data set WORK.MAILED has 4 observations and 8 variables.  
NOTE: The DATA statement used 0.48 seconds.

```

50
51
52  proc sort data=mailed;
53  by pwcid;
54  run;

```

NOTE: The data set WORK.MAILED has 4 observations and 8 variables.  
NOTE: The PROCEDURE SORT used 0.05 seconds.

```

55
56
57
58  *****/
59  merge samples dataset with mailed dataset
60  keep only records from mailed dataset
61  rename variables so that names are consistent with those in screener sample ASCII file
62  *****/
63  data mailed (rename=(address=custadd city=custcity state=cusstate zip=custzip company=custname
64  contact=custcont phone1=custphon));
65  merge samples (in=a) mailed (in=b);
66  by pwcid;
67  if b;
68  run;

```

NOTE: The data set WORK.MAILED has 4 observations and 43 variables.  
NOTE: The DATA statement used 0.05 seconds.

```

69
70
71
72  *****/
73  create permanent data set containing records from most recent mailout
74  each time a new batch is created, the usps.main sample dataset will end in a
75  different number (e.g.,usps.main sample1, usps.main sample2)
76  *****/
77  data main.sample&i;
78  set mailed;
79  run;

```

NOTE: The data set MAIN.SAMPLE8 has 4 observations and 43 variables.  
NOTE: The DATA statement used 1.1 seconds.

```

80
81
82
83
84  *****/
85  output CATI sample to fixed width ASCII file
86  each time a new ASCII is created, the usps main sample.asc file will end in a
87  different number (e.g.,main.main sample1.asc, main.main sample2.asc)
88  *****/
89  data main.sample&i;
90  set main.sample&i;
91  file "s:\ogs\common\usps\main study\questionnaire\questionnaire sample\usps main sample&i..asc" lrecl=360;
92  put @1 custphon $10. /**permit-owner's phone number**/
93  @15 pwcid 5. /**identifying code**/
94  @20 custcont $20. /**company that owns permit**/
95  @45 custadd $25. /**company address**/
96  @75 custcity $20. /**company city**/
97  @100 cusstate $2. /**company state**/

```

```

98      @105 custzip $5.      /**company zip**/
99      @115 custname $50.    /**name of contact person at company**/
100     @170 permnum $5.      /**FCM permit number**/
101     @180 permtype $29.    /**type of mail: MT, PC or PI**/
102     @210 permdate mmddyy8. /**date that permit was issued**/
103     @220 totbulk comma15. /**total bulk volume for 1998**/
104     @240 cat1vol comma15. /**rate group 1 volume for 1998**/
105     @260 cat2vol comma15. /**rate group 2 volume for 1998**/
106     @280 cat3vol comma15. /**rate group 3 volume for 1998**/
107     @300 strata 1.        /**1=small 2=medium 3=large 4=certainty**/
108     @305 pocity $22.      /**city of issuing post office**/
109     @335 postate $2.      /**state of issuing post office**/
110     @340 pozip $5.        /**ZIP Code of issuing post office**/
111     @350 count 1.        /**number of permits this company has in the sample**/
112     @355 called 1.;      /**1=company was called in pretest about a different permit**/
113     run;

```

NOTE: The file "s:\ogs\common\usps\main study\questionnaire\questionnaire sample\usps main sample8.asc" is:  
FILENAME=s:\ogs\common\usps\main study\questionnaire\questionnaire sample\usps main sample8.asc,  
RECFM=V,LRECL=360

NOTE: 4 records were written to the file "s:\ogs\common\usps\main study\questionnaire\questionnaire sample\usps main sample8.asc".  
The minimum record length was 355.  
The maximum record length was 355.

NOTE: The data set MAIN.SAMPLE8 has 4 observations and 43 variables.

NOTE: The DATA statement used 2.14 seconds.