The Public Representative hereby replies to the reply comments the Postal Service submitted in this docket.¹ The Postal Service responds to several concerns and/or criticisms made by the Public Representative in his initial comments.² Specifically, the Postal Service responds to four issues:

(1) The Public Representative’s concern that the parameters of the variables used to determine volume variability, estimated by full quadratic specifications, could not be tested as to whether or not they were jointly significant from zero. USPS Reply Comments at 2-5.

¹ Reply Comments Of The United States Postal Service To The Initial Comments Of The Public Representative, August 27, 2019 (USPS Reply Comments).

² Public Representative Comments On Proposed Change In Periodic Reporting, August 21, 2019 (PR Comments).
(2) The Public Representative’s criticism that the Postal Service eliminated legitimate observations from its CPMS dataset in a way which resulted in biased estimates of collection variability. *Id.* at 6-8.

(3) The Public Representative’s criticism that the Postal Service should not have moved peak season LDC 24 hours incurred by full-time carriers from the Sunday Delivery cost pool to the Regular Delivery cost pool. *Id.* at 8-9.

(4) The Public Representative’s criticism that the Postal Service should not have moved *moved SPR relay hours to regular city carrier time pools*. *Id.* at 9-10.

The Public Representative will reply to each of the Postal Service’s responses in the following text.

*Inability To Test For Joint Significance*

The Postal Service responds to the Public Representative’s concern that a test of the joint significance of the parameters which are used to calculate volume variability in the Postal Service’s models cannot be performed. The Public Representative was concerned that the Full Quadratic models, as specified by the Postal Service, contained too many variables which were similar to each other—particularly volume and volume cross-product terms. The inability to perform this test suggested the possibility of extreme multicollinearity.

The Postal Service explains that:

“When a righthand-side variable takes on very large values relative to the dependent variable the resulting estimated coefficient is often very small in absolute value. This result happens in the delivery regressions because of the very large value for the square of volume, relative to the SPR delivery hours.”

USPS Reply Comments at 2. This can result in a standard error of the squared volume term so low that SAS is unable to perform the test of joint significance, rather than representing extreme multicollinearity. *Id.* at 2-3.

The Postal Service explains that that an easy solution to this problem is to scale the volume observations up by a factor of 1000. By scaling volume in this manner
the Postal Service shows that the parameter estimates remain unaltered, and the parameters used to calculate volume variability are jointly significant. *Id.* at 2-5.

The Public Representative accepts the Postal Service’s explanation. The Full Quadratic specification does yield volume parameter estimates that are jointly significantly different from zero. The Public Representative remains concerned that this specification still contains a significant amount of multicollinearity, which may result in non-operationally based changes in volume variability in subsequent estimates. In particular, the Public Representative is concerned that there is a very high level of correlation between volume and volume cross products.

This becomes more evident when considering that there are three sets of dummy variables, each with several terms, which are multiplied by or “crossed with” volume: (1) the type of route: curbline, central, etc., (route-type); (2) business versus residential; and (3) activity at the receptacle, e.g., delivery in or at the mailbox, delivery at the front door, etc., (receptacle activity). Because the Postal Service constructs these dummy variables as percentages of volume, their volume may be summed or combined within each type of dummy. If one the runs the full quadratic models after this summation, the extent of correlation between volume and its cross-products becomes more clearly visible. The attached SAS output shows that the delivery route (or delivery technology) terms are have a 99 percent correlation with volume, and volume has a negative 59 percent correlation with volume squared.³

The Public Representative recognizes that the Commission has accepted the Full Quadratic Model (so long as it produces significant and meaningful results). Stability of estimates over time has not been raised as an issue before. For that reason, the Public Representative accepts the Postal Service’s use of the Full Quadratic Model in this proceeding, as specified, but recommends the Commission consider further

³ See, Attachment 1. The Public Representative only had time to test this notion on the December Regular Delivery model, but expects similar results will occur in the other Full Quadratic models with this specification.
investigation of the issue of variability instability over time due to very high multicollinearity when dummy variables of the variable of interest are created as a percent of volume

**Excluding Non-Erroneous Collection Observations**

The Public Representative argued that the Postal Service determined that an average collection box could not hold more than approximately 825 pieces of mail, and eliminated observations from collection boxes which contained more than 825 pieces of mail, and did not “provide evidence that a box with more mailpieces than 825 is overflowing and incapable of holding more mail.” PR Comments, 22,

The Postal Service’s Reply Comments clarify that it did not eliminate observations based on whether an individual box held more than 825 pieces a single day. Rather, it excluded observations where the average number of pieces for all boxes in a finance number were greater than 825. USPS Comments, 73. The Postal Service considers this measure to be a non-arbitrary, operationally based, method. It states, “by using the average overall boxes, this standard does allow for the possibility that on a given day a finance number may be required to sweep one or two “overstuffed” boxes, but eliminates observations with clear data problems. It is extremely unlikely that all of the boxes within a finance number’s area would be overstuffed on the same day. USPS Comments, 73.

The Public Representative remains unconvinced that this method is non-arbitrary simply because it compares the average daily volume of all boxes in a finance number to an operationally based maximum value. The problem is two-fold. First, the operational maximum does not identify the type of collection box used to determine the maximum number of pieces able to fit in a collection box. Data provided by the Postal Service show that average daily collection volumes across all finance numbers taken
from Jumbo Snorkel and Large Snorkel collection boxes, exceeds 825 pieces. If so, the Postal Service’s claim that “[i]t is extremely unlikely that all of the boxes within a finance number’s area would be overstuffed (i.e. greater than 825) on the same day” does not hold up. Id., 73

Second, the Postal Service verified that collection boxes with high daily collection volumes primarily occur in less than one-half of total finance numbers in the sample, and “[t]hree-fourths of the finance numbers in this group have only one or two high volume days.” USPS Comments, 71-72. While these high, daily collection volumes may not be typical, they actually occur. Since the data show that two types of collection boxes have average daily volumes across all finance numbers greater than 825, they confirm these are potentially legitimate observations.

The Postal Service does state that “[w]ith a few exceptions, these high collection volume days are not typical of the recorded finance numbers’ normal operations. The 160 finance number days in the top one percent occur at 76 different finance numbers. Three-fourths of the finance numbers in this group have only one or two high volume days. The Public Representative agrees the days and locations are not typical, but disagrees that all atypical observations represent problems related to collection boxes being “overstuffed,” especially as Table 1 shows, nearly 70 percent of Saturday, and 59 percent of sample observations would have been deleted if missing data points had been deleted. Keeping these observations in the sample biases the calculation of daily average collection volumes greatly downward.

This clearly presents a problem, since it is unlikely the Postal Service has another collection volume data source, or that it can reconstruct the original data scans (if scans were even made) and determine whether scan errors which result in missing data points can be reconstructed to reveal a positive collection volume. If the Postal

———

4 See, Attachment 2, the text version of the file entitled “Reply Boxtype Collection NP.sas.” The PR does not include the output, since the data file is confidential. Simply paste the text into a blank SAS file, change the path to the data source and run the program.
Service does not have another useable collection mail dataset with collection volumes and associated time, or if it cannot correct the missing data points, the Commission has few choices. It can either accept the Postal Service’s final model, which relies upon removing observations where the average daily collection volume exceed 825 pieces in any finance number, or accept the Postal Service’s initial method of dropping the observations with the highest one percent of daily top one percent of daily collection volumes, the results of which are presented in Table 31 of the Postal Service’s Report submitted with its Comments.\(^5\)

**Missing Collection Observations**

The Public Representative has another concern after being provided more time to examine the collection box programs. The Postal Service turns all missing data points and all data points recorded with an ‘X’ to be zero. However, it does not follow that all of these failures would have been zero if the carrier had properly recorded them. The converse is more likely to be the case. The Public Representative ran “Reply Boxtype Collection NP.sas” in Attachment 2 by deleting the observations with these missing data points, and found that. Table 1 presents the number of observations that either are recorded incorrectly with an ‘X’, or are missing.\(^6\)

---

\(^5\) See, Petition of the United States Postal Service For the Initiation of a Proceeding to Consider Proposed Changes in Analytical Principles (Proposal One), “A New Study of Special Purpose Route Carrier Costs” by Professor Michael D. Bradley, filed June 21, 2019

\(^6\) The Public Representative believes that data points recorded with an ‘X’ also reflect instances where the carrier did not record the volume picked up from a collection box. A manager, or other Postal Service employee, probably inserted an ‘X’ sometime after the carrier collection data was electronically saved.
Table 1
Deleted Collection Observations

<table>
<thead>
<tr>
<th># of Observations</th>
<th>M-F</th>
<th>Saturday</th>
<th>Total</th>
<th>Number of Obs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original</td>
<td>199</td>
<td>408,712</td>
<td>608</td>
<td>Original</td>
</tr>
<tr>
<td>After Deletion</td>
<td>142</td>
<td>125,406</td>
<td>267</td>
<td>After Deletion</td>
</tr>
<tr>
<td>Deleted</td>
<td>56</td>
<td>283,306</td>
<td>340</td>
<td>Deleted</td>
</tr>
<tr>
<td>% Deleted</td>
<td>28.5%</td>
<td>69.3%</td>
<td>55.9%</td>
<td>% Deleted</td>
</tr>
</tbody>
</table>

Source: Attachment 2 with code deleting ‘X’ and ‘.’ and not making them zero values.

The data show that 28.5 percent of collection mail observations in the Monday through Friday sample, and 69.3 percent of collection mail observations in the Saturday sample were missing. While one would expect carriers to be less diligent about recording collection mail on Saturday, it is surprising that nearly 70 percent of the Saturday sample had missing observations, since the data were collected as part of an annual study, where greater managerial supervision should have been present. This result is troubling, since it calls into question the representativeness of the collection sample, especially Saturday, which begins with double the observations compared to the Monday through Friday sample.

The Public Representative recommends the Commission ask the Postal Service to examine whether there is another, more reliable, dataset of collection volume, time and other causal factors that it might use, and/or whether the existing dataset was based on electronic scans made by the carrier which can be corrected. If neither of these options exist, the Commission will be left with a decision to either reject the proposed collection model; accept the final model based on the practice of deleting observations considered to be “overstuffed;” or accept the model where the top one percent of observations are deleted. See, Table 31 of the Proposed Report. Proposed Report, 72.
The Public Representative recommends the Commission not reject the collection model, since it would probably be required to reject the entire study. The Postal Service rejects dropping the top one percent of collection volume observations, because “there is no reliable basis for determining which observations should be excluded.” Proposed Report, 72. Rather, it proposes an operationally based method of deleting questionable observations.

The Public Representative’s Reply Comments support his conclusion that the operational, overstuffed method sets a threshold too low and does not take into account the size of different collection boxes. Simply increasing the threshold to the average value of the highest volume type of collection box, fails to account for the reduction of average daily collection volume per box in any finance number due to collection box types that can hold very little volume. Consequently, the Public Representative proposes the Commission require the Postal Service to use the “leverage” method and subsequent investigation followed by the Postal Service in the Monday through Saturday, Regular Delivery Model. Id., 59-65.

Moving Sunday Hours Delivered By Full-Time Carriers To The Monday – Saturday Sample

The Postal Service says the total volume delivered by full-time carriers were sorted and prepared on Saturday at their normal Monday through Friday destination delivery unit (DDU), and are simply picked up at the DDU on Sunday and delivered by full-time carriers. Id., 69. The implication is that office time should be associated with delivery time. However, if part of the office time that full-time carriers incur on a Saturday is associated with total volume delivered on Sunday, there will be a large mismatch of time and volume for full-time regular delivery carriers.

However, the Postal Service does not address the question “how much time full-time carriers spend in the office on Sunday handling this already prepared mail,” nor does it address the question “whether additional parcel volumes arrive at the DDU on Sunday before full-time carriers arrive.” If these questions show a positive time, full-time carriers do incur office time preparing parcels that are not overflow mail. They also incur
some office time handling the already prepped mail on Saturday, which must be
delivered on Sunday. Full-time carriers who deliver parcels on Sunday incur office time
handling and loading parcels that were sorted on Saturday, and they incur office time
sorting, prepping, handling and loading parcels that arrive after they clock out on
Saturday and before they clock in on Sunday.

The Public Representative is not strongly opposed to treating Sunday hours
incurred by full-time city carriers as belonging to the Monday – Saturday cost pool.
However, the Postal Service should provide the Commission with data on the amount of
office time incurred by carriers prepping and loading overflow mail on Sunday, the time
they incur sorting, prepping and loading mail which arrives after they clock out of the
office on Saturday and before they clock into the office on Sunday, and the time full-time
carriers spend prepping the volume of mail which turns out to be “overflow to Sunday”
mail.

With these data, the Commission should be able to determine whether office time
incurred on Sunday is or is not approximately the same as the office time incurred
handling the parcels on Saturday that become overflow mail. If they are approximately
the same, it would be more appropriate to retain full-time carrier hours incurred in the
office and on delivery on Sunday as part of the Sunday delivery cost pool, since the
Postal Service includes Sunday office and delivery time for full-time city carriers as a
component of Monday through Saturday accrued costs, but does not include this time or
volume in its Regular Delivery Models. For this reason, it would be more accurate to
retain the Sunday delivery time and volume delivered by full-time carriers as part of the
Sunday Delivery model.

Relay

The Postal Service stated in its Proposed Report “[b]ecause these hours are
caused by mail delivered on letter routes and not by SPR delivered volumes; they are
appropriately attributed to letter route costs.” Proposed Report, 75. The Public
Representative criticized this explanation as a product of reductionist logic in his
Comments. PR Comments, 12. The Postal Service defends its logic by referring back to the Proposed Report. USPS Reply Comments, 10. The Postal also correctly states that moving SPR relay hours to city carrier delivery is an accepted methodology. Proposed Report, fn. 31, 7; and USPS Reply Comments, 9. However, it merely mentions this, but provides no citation to support its claim.

The Public Representative reviewed various documents and files, including the 2018 Summary Description of Segments and Components (2018 Summary Description), the Preface to CCCS-SPR (CCCS-SPR Preface), and the B-workpapers.\(^7\) The 2018 Summary Description does not mention this transfer. Rather, it states that Delivery Support, Component 280 “contains the activities associated with driving to and from routes and relays on both letter routes and special purpose routes.”\(^8\) The CCCS-SPR Preface states “it collects data on “all Relay routes and Other routes with low work hour usage.” CCCS-SPR Preface, 17. Only the segment 7 Excel file in the B-Workpapers cites the accepted relay cost methodology to which the Postal Service refers, although the reference simply states “Relay (moved to letter routes in WS 7041 l6). Unfortunately, this workbook is not available in electronic form accessible by Commission staff.\(^9\) Nevertheless, the Public Representative agrees moving SPR relay costs to regular delivery is the Commission’s accepted methodology.

---


\(^8\) 2018 Summary and Description, file “CS07-18.docx, 6.

\(^9\) See, 2018 B-Workpapers, CS06&7-Public-FY18.xlsx, tab 7.0.5, cell B37.
Conclusion

The Public Representative accepts the transfer of SPR relay costs to regular city carrier delivery. It also proposes the Commission accept its recommendations with regard to further examination of multicollinearity among the volume variable, its square and its cross-terms; further examination of the method of removing collection volume outliers; and examination of the number of office hours regular delivery carriers spend on Saturday associated with the overflow mail delivered on Sunday, compared to the office hours spent by full-time carriers in the office on Sundays handling and loading mail already prepped, as well as the hours associated with sorting, prepping and loading mail which has arrived before the carrier arrives at the DDU on Sunday.

Respectfully submitted,

Lawrence Fenster
Public Representative

901 New York Avenue, N.W., Suite 200
Washington, DC 20268-0001
Phone (202) 789-6862
Email: larry.fenster@prc.gov
ATTACHMENT 1

SAS Output Showing Correlation Among Volume Terms

<table>
<thead>
<tr>
<th>Variable</th>
<th>Intercept</th>
<th>Vol</th>
<th>vol2</th>
<th>voldeltech</th>
<th>volbr</th>
<th>voldelmode</th>
</tr>
</thead>
<tbody>
<tr>
<td>vol</td>
<td>-0.3%</td>
<td>100.0%</td>
<td>-59.0%</td>
<td>-99.2%</td>
<td>-45.9%</td>
<td>9.6%</td>
</tr>
</tbody>
</table>

SAS Program Text For Showing Correlation Among Volume Terms

libname SPR 'Z:\RM2019-6\USPS\Public\USPS-RM2019-6_1\SPR.Prop.1.Fldr.1.Public Files\Public Folder\Directory 1 Analysis Data Sets';

*******************SEPT JT CONFIDENCE TEST *******************;
options nodate;
ods graphics off;

*** Reading in the four regression data sets ***;

data sepreg1; set spr.sepreg5;
data decreg1; set spr.decreg5;
data junreg1; set spr.junreg5;
data marreg1; set spr.marreg5;

**** Selecting the month to be estimated ***;
data sepreg2;
set sepreg1;

**** Selecting the regular delivery observations ***;
**** Selectin the SPR locations size *****;
**** Creating a month index *****;

data sepreg3; set sepreg2 ;
if ldc = ’2300’ or ldc = ’2400’;
if s_thours ge 96;
if DEC = ’.’ then dec = 0;
if SEP = ’.’ then sep = 0;
if JUN = ’.’ then jun = 0;
if MAR = ’.’ then MAR = 0;
if vol > 0;

*** Converting the characteristic variables to levels ***********;
*** for calculating ratios for combined LDC23 and 24 observations *****;
data sepreg3; set sepreg3;
cbuL = cbu * DT ;
centL = cent * DT ;
curbL = curb * DT ;
doorL = door * DT ;
otherL = other * DT ;
FDPl = FDP * DE ;
IAML = IAM * DE ;
IFRL = IFR * DE ;
ODEl = ODE * DE ;
BRL = BR * DT ;

**** Cumulating data by finance number / date ****;
proc sort data=sepreg3; by fin_no date;
proc summary data=sepreg3; by fin_no date;
Var
cbuL
centL
curbL
doorL
otherL
FDPl
IAML
IFRL
ODEl
BRL
DT
DE
hours
ndhours
thours
vol
boxes;
output out=sepreg3a sum=
cbuL
centL
curbL
doorL
otherL
FDPI
IAMI
IFRI
ODEI
brL
DT
DE
hours
ndhours
thours
vol
s_boxes
mean=
a_cbuL
a_centL
a_curbL
a_doorL
a_otherL
a_FDPL
a_IAML
a_IFRL
a_ODEL
a_BRL
a_DT
a_DE
a_hours
a_ndhours
a_thours
a_vol
boxes;

;

*** Calculating characteristic variables for combined LDCs ***;
data sepreg3a; set sepreg3a;
cbu =   cbuL /  DT  ;
cent =   centL /  DT  ;
curb =   curbL /  DT  ;
door =   doorL /  DT  ;
other =   otherL/  DT  ;
FDP = FDPL / DE ;
IAM = IAML / DE ;
IFR = IFRL / DE ;
ODE = ODEL / DE ;
br = BRL / DT ;

DATA sepreg3B; set sepreg3a;
keep fin_no date
cbu
cent
curb
door
other
FDP
IAM
IFR
ODE
br
DT
hours
ndhours
thours
vol
boxes;
run;

*** Constructing variables needed for econometric equation ***;

data sepreg3C; set sepreg3B;
cv=boxes;
cv2 = cv * cv;
vol2=vol*vol;
curb2 = curb * curb ;
cbu2 = cbu * cbu ;
cent2 = cent * cent ;
door2 = door * door ;
br2 = br * br ;
iam2 = iam * iam ;
fdp2 = fdp * fdp ;
ifr2 = ifr * ifr ;
volcurb = vol * curb;
volcbu = vol * cbu;
volcent = vol * cent;
voldoor = vol * door;
volbr = vol * br;
voliam = vol * iam;
volfdp = vol * fdp;
volfir = vol * ifr;
volode = vol * ode;
volcv = vol * cv;

cvcurb = cv * curb;
cvcbu = cv * cbu;
cvcnt = cv * cent;
cvdoor = cv * door;
cvbr = cv * br;
cviam = cv * iam;
cvfdp = cv * fdp;
cvifr = cv * ifr;
cvode = cv * ode;

curbcbu = curb * cbu;
curbcent = curb * cent;
curbdoor = curb * door;
curbbbr = curb * br;
curbiam = curb * iam;
curbfdfp = curb * fdp;
curbifr = curb * ifr;

cbucbnt = cbu * cent;
cbudoor = cbu * door;
cbubr = cbu * br;
cbuiam = cbu * iam;
cbufdp = cbu * fdp;
cbuifr = cbu * ifr;

centdoor = cent * door;
centr = cent * br;
centiam = cent * iam;
centfdp = cent * fdp;
centifr = cent * ifr;
doorbr = door * br ;
dooriam = door * iam ;
doorfdp = door * fdp ;
doorifr = door * ifr ;

briam = br * iam ;
brfdp = br * fdp ;
brifr = br * ifr ;

lamfdp = lam * fdp ;
lamifr = lam * ifr ;

fdpifr = fdp * ifr ;

data sepreg4; set sepreg3C;
newvol = sum(vol,volcurb,volcbu,volcent,voldoor,volbr,voliam,volfdb,volifr,volode,volcv);
newvol2 = newvol*newvol;
run;

data testvolreg;
   set sepreg4 (keep= thours newvol newvol2 cv cv2 cvcurb cvcbu cvcent cvdoor
cvbr cviam cvfdp cvifr
curb cbu cent door br iam fdp ifr curb2 cbu2 cent2 door2 br2 iam2 fdp2 ifr2);
run;

proc reg data = testvolreg;
model thours=newvol newvol2
cv cv2 cvcurb cvcbu cvcent cvdoor
cvbr cviam cvfdp cvifr curb cbu cent
door br iam fdp ifr curb2 cbu2 cent2 door2 br2 iam2 fdp2 ifr2
    /vif white collin ;

proc corr data = testvolreg;
run;
Attachment 2  
SAS Program Showing Collection Volume Statistics of Different Collection Boxes

libname SPR 'Z:\RM2019-6\USPS\NPLR\Directory 1 - Operational Data\CPMS Data\';

Data mf;
set spr.all_points_m_f;
boxtype = BOX_TYPE_CODE;

Data mf1; set mf;
if   boxtyp e eq 'EXPR SNKL'
or   boxtyp e eq 'EXPRESS'
or   boxtyp e eq 'JUMBO SNKL'
or   boxtyp e eq 'JUMBO STD'
or   boxtyp e eq 'LARGE'
or   boxtyp e eq 'LARGE SNKL'
or   boxtyp e eq 'PRIORITY'
or   boxtyp e eq 'PRTY SNKL'
or   boxtyp e eq 'SNORKEL'
or   boxtyp e eq 'STANDARD'
;

***** SET 'X' TO ZERO *****;
if   Sep_182017 = 'X' then Sep_182017 = 0;
if   Sep_192017 = 'X' then Sep_192017 = 0;
if   Sep_202017 = 'X' then Sep_202017 = 0;
if   Sep_212017 = 'X' then Sep_212017 = 0;
if   Sep_222017 = 'X' then Sep_222017 = 0;
if   Sep_252017 = 'X' then Sep_252017 = 0;
if   Sep_262017 = 'X' then Sep_262017 = 0;
if   Sep_272017 = 'X' then Sep_272017 = 0;
if   Sep_282017 = 'X' then Sep_282017 = 0;
if   Sep_292017 = 'X' then Sep_292017 = 0;

ZIP = substr(LOCATION_ID,1,5);
length box_nbr $5;
BOX_NBR = substr(LOCATION_ID,6,5);

*** Convert Character Values to Numbers  *8*;
data mf2; set mf1;
sep_182017N=input(sep_182017, 8.);
drop sep_182017;
rename sep_182017N = sep_182017;

sep_192017N=input(sep_192017, 8.);
drop sep_192017;
rename sep_192017N = sep_192017;

sep_202017N=input(sep_202017, 8.);
drop sep_202017;
rename sep_202017N = sep_202017;

sep_212017N=input(sep_212017, 8.);
drop sep_212017;
rename sep_212017N = sep_212017;

sep_222017N=input(sep_222017, 8.);
drop sep_222017;
rename sep_222017N = sep_222017;

sep_252017N=input(sep_252017, 8.);
drop sep_252017;
rename sep_252017N = sep_252017;

sep_262017N=input(sep_262017, 8.);
drop sep_262017;
rename sep_262017N = sep_262017;

sep_272017N=input(sep_272017, 8.);
drop sep_272017;
rename sep_272017N = sep_272017;

sep_282017N=input(sep_282017, 8.);
drop sep_282017;
rename sep_282017N = sep_282017;

sep_292017N=input(sep_292017, 8.);
drop sep_292017;
rename sep_292017N = sep_292017;
proc contents;
*** Read In Saturday CPMS Data ***;
*** Identify Collection Points Served by SPR Carriers ***;
*** Create Box Number Variable ***;

data sat1; set spr.all_points_sat;
boxtyp = BOX_TYPE_CODE;
data sat2; set sat1;

if boxtyp eq 'EXPR SNKL'
or boxtyp eq 'EXPRESS'
or boxtyp eq 'JUMBO SNKL'
or boxtyp eq 'JUMBO STD'
or boxtyp eq 'LARGE'
or boxtyp eq 'LARGE SNKL'
or boxtyp eq 'PRIORITY'
or boxtyp eq 'PRTY SNKL'
or boxtyp eq 'SNORKEL'
or boxtyp eq 'STANDARD';

ZIP = substr(LOCATION_ID,1,5);
length box_nbr $5;
BOX_NBR = substr(LOCATION_ID,6,5);

proc contents;
proc sort data=mf2; by zip box_nbr;
proc sort data=sat2; by zip box_nbr;

*** Merge M-F Data and Saturday Day by ZIP and Box Number ***;
data alldays; merge mf2(merge w) sat2(merge s) by zip box_nbr;
source='CPMS';

if w = 1 and s = 1 then source = 'bth';
if w = 1 and s = 0 then source = 'jsw';
if w = 0 and s = 1 then source = 'jss';
if \( w = 0 \) and \( s = 0 \) then source = 'non';

*** Check Merge ***;
data both; set alldays;
if source eq 'bth';
proc contents;

data Just_week; set alldays;
if source eq 'jsw';
data just_sat; set alldays;
if source eq 'jss';
data none; set alldays;
if source eq 'non';

*** Set any missing values to zero***;
data alldays1; set alldays;
if Sep_092017 = '.' then Sep_092017 = 0;
if Sep_162017 = '.' then Sep_162017 = 0;
if Sep_182017 = '.' then Sep_182017 = 0;
if Sep_192017 = '.' then Sep_192017 = 0;
if Sep_202017 = '.' then Sep_202017 = 0;
if Sep_212017 = '.' then Sep_212017 = 0;
if Sep_222017 = '.' then Sep_222017 = 0;
if Sep_232017 = '.' then Sep_232017 = 0;
if Sep_252017 = '.' then Sep_252017 = 0;
if Sep_262017 = '.' then Sep_262017 = 0;
if Sep_272017 = '.' then Sep_272017 = 0;
if Sep_282017 = '.' then Sep_282017 = 0;
if Sep_292017 = '.' then Sep_292017 = 0;
if Sep_302017 = '.' then Sep_302017 = 0;

sep1 = Sep_092017;
sep2 = Sep_162017;
sep3 = Sep_182017;
sep4 = Sep_192017;
sep5 = Sep_202017;
sep6 = Sep_212017 ;
sep7 = Sep_222017 ;
sep8 = Sep_232017 ;
sep9 = Sep_252017 ;
sep10 = Sep_262017 ;
sep11 = Sep_272017 ;
sep12 = Sep_282017 ;
sep13 = Sep_292017 ;
sep14 = Sep_302017 ;

*** Sum volume across days***;
data sum; set alldays1;
sum_range=sum(of sep1-sep14);
totvol = sum_range;

DATA unitbox; set sum;

proc sort  SORTSEQ=linguistic
  (NUMERIC_COLLATION=ON);
  by boxtype;

data avgday; set unitbox;
  count = 14;
  if totvol = 0 then unitvol=0;
  unitvol = totvol/count;

proc univariate data = avgday; by boxtype;
  var unitvol;
  title1 'Data By Boxtype';
  run;
  quit;
libname SPR 'Z:\RM2019-6\USPS\NPLR\Directory 1 - Operational Data\CPMS Data\';

Data mf;
set spr.all_points_m_f;
boxtype = BOX_TYPE_CODE;

Data mf1; set mf;

if  boxtype   eq 'EXPR SNKL'
or  boxtype   eq 'EXPRESS'
or  boxtype   eq 'JUMBO SNKL'
or  boxtype   eq 'JUMBO STD'
or  boxtype   eq 'LARGE'
or  boxtype   eq 'LARGE SNKL'
or  boxtype   eq 'PRIORITY'
or  boxtype   eq 'PRTY SNKL'
or  boxtype   eq 'SNORKEL'
or  boxtype   eq 'STANDARD'
;

**** SET 'X' TO DELETE *****;

if      Sep_182017 = 'X' then delete;
if      Sep_192017 = 'X' then delete;
if      Sep_202017 = 'X' then delete;
if      Sep_212017 = 'X' then delete;
if      Sep_222017 = 'X' then delete;
if      Sep_252017 = 'X' then delete;
if      Sep_262017 = 'X' then delete;
if      Sep_272017 = 'X' then delete;
if      Sep_282017 = 'X' then delete;
if      Sep_292017 = 'X' then delete;

ZIP = substr(LOCATION_ID,1,5);
length box_nbr $5;
BOX_NBR = substr(LOCATION_ID,6,5);

*** Convert Character Values to Numbers *8*;
data mf2; set mf1;
sep_182017N=input(sep_182017, 8.);
drop sep_182017;
rename  sep_182017N = sep_182017;

sep_192017N=input(sep_192017, 8.);
drop sep_192017;
rename  sep_192017N = sep_192017;

sep_202017N=input(sep_202017, 8.);
drop sep_202017;
rename  sep_202017N = sep_202017;

sep_212017N=input(sep_212017, 8.);
drop sep_212017;
rename  sep_212017N = sep_212017;

sep_222017N=input(sep_222017, 8.);
drop sep_222017;
rename  sep_222017N = sep_222017;

sep_252017N=input(sep_252017, 8.);
drop sep_252017;
rename  sep_252017N = sep_252017;

sep_262017N=input(sep_262017, 8.);
drop sep_262017;
rename  sep_262017N = sep_262017;

sep_272017N=input(sep_272017, 8.);
drop sep_272017;
rename  sep_272017N = sep_272017;

sep_282017N=input(sep_282017, 8.);
drop sep_282017;
rename  sep_282017N = sep_282017;

sep_292017N=input(sep_292017, 8.);
drop sep_292017;
rename  sep_292017N = sep_292017;
proc contents;

*** Read In Saturday CPMS Data **;
*** Identify Collection Points Served by SPR Carriers ***;
*** Create Box Number Variable ***;

data sat1; set spr.all_points_sat;
boxtype = BOX_TYPE_CODE;

data sat2; set sat1;

********** DELETE MISSING OBSERVATIONS **********;

if Sep_092017 = '.' then delete;
if Sep_162017 = '.' then delete;
if Sep_232017 = '.' then delete;
if Sep_302017 = '.' then delete;

if boxtype eq 'EXPR SNKL'
or boxtype eq 'EXPRESS'
or boxtype eq 'JUMBO SNKL'
or boxtype eq 'JUMBO STD'
or boxtype eq 'LARGE'
or boxtype eq 'LARGE SNKL'
or boxtype eq 'PRIORITY'
or boxtype eq 'PRTY SNKL'
or boxtype eq 'SNORKEL'
or boxtype eq 'STANDARD'

ZIP = substr(LOCATION_ID,1,5);
length box_nbr $5;
BOX_NBR = substr(LOCATION_ID,6,5);

proc contents;

proc sort data=mf2; by zip box_nbr;
proc sort data=sat2; by zip box_nbr;

*** Merge M-F Data and Saturday Day by ZIP and Box Number ***,
data alldays; merge mf2(in=w) sat2(in=s) ; by zip box_nbr;

source='CPMS';

if w = 1 and s = 1 then source = 'bth';
if w = 1 and s = 0 then source = 'jsw';
if w = 0 and s = 1 then source = 'jss';
if w = 0 and s = 0 then source = 'non';

*** Check Merge ***;
data both; set alldays;
if source eq 'bth';

proc contents;

data Just_week; set alldays;
if source eq 'jsw';

data just_sat; set alldays;
if source eq 'jss';

data none; set alldays;
if source eq 'non';

data alldays1; set alldays;

sep1 = Sep_092017 ;
sep2 = Sep_162017 ;
sep3 = Sep_182017 ;
sep4 = Sep_192017 ;
sep5 = Sep_202017 ;
sep6 = Sep_212017 ;
sep7 = Sep_222017 ;
sep8 = Sep_232017 ;
sep9 = Sep_252017 ;
sep10 = Sep_262017 ;
sep11 = Sep_272017 ;
sep12 = Sep_282017 ;
sep13 = Sep_292017 ;
sep14 = Sep_302017;

*** Sum volume across days***;
data sum; set alldays1;
sum_range=sum(of sep1-sep14);
totvol = sum_range;

DATA unitbox; set sum;

proc sort  SORTSEQ=linguistic
   (NUMERIC_COLLATION=ON);
   by boxtype;

data avgday; set unitbox;
count = 14;
if totvol = 0 then unitvol=0;
unitvol = totvol/count;

proc univariate data = avgday; by boxtype;
var unitvol;
title1 'Data By Boxtype';
run;
quit;