

UNITED STATES OF AMERICA
Before The
POSTAL RATE COMMISSION
WASHINGTON, D.C. 20268-0001

Postal Rate and Fee Changes, 2006)

Docket No. R2006-1

RESPONSES OF OFFICE OF THE CONSUMER ADVOCATE
WITNESS J. EDWARD SMITH TO INTERROGATORIES OF
ADVO (ADVO/OCA-T3-25-32)
(October 10, 2006)

The Office of Consumer Advocate hereby submits responses of J. Edward Smith to interrogatories ADVO/OCA-T3-25-32, dated September 26, 2006. Each interrogatory is stated verbatim and is followed by the response.

Respectfully submitted,

SHELLEY S. DREIFUSS
Director
Office of Consumer Advocate

Emmett R. Costich
Attorney

901 New York Avenue, NW Suite 200
Washington, D.C. 20268-0001
(202) 789-6830; Fax (202) 789-6891
e-mail: rand.e.costich@prc.gov

RESPONSES OF OCA WITNESS J. EDWARD SMITH
TO INTERROGATORIES ADVO/OCA-T3-25-32

ADVO/OCA-T3-25. Please provide the SAS log for ReadVolume SAS program in OCA LR L-4, Section 2.

RESPONSE TO ADVO/OCA-T3-25.

Please see my answer to USPS/OCA-T3-10.

RESPONSES OF OCA WITNESS J. EDWARD SMITH
TO INTERROGATORIES ADVO/OCA-T3-25-32

ADVO/OCA-T3-26. Please refer to OCA LR L-4, Section 2, ReadVolume SAS program, on the first page under "Deleting missing time data and setting missing data for a number of variables equal to zero." For observations in which the following variables contained missing data, values for these variables were set to zero. The variables are: autoflats, autoltrs, casflts,casltrs, DPS, miles, prcl, pri, seqflts, and seqltrs.

(a) Please explain your rationale behind setting values for these missing data equal to zero.

(b) Did you attempt to check this assumption with the individuals at the USPS who are responsible for the DOIS data collection and database? Please explain.

(c) Do you consider setting the missing data values equal to zero less arbitrary than setting these values equal to other values, such as one, five, ten, etc.?

(d) Since true values for variables with missing values are unknown, please explain how your treatment is better than deleting observations with missing values from the database.

(e) Please confirm that if missing data observations contain no new information relative to that contained in non-missing data, then parameter estimates for variables using non-missing data remain unbiased. If you cannot confirm, please explain fully.

(f) Please confirm that the loss of efficiency from higher variances for parameter estimates formed from deleting observations decreases as sample sizes get larger. If you cannot confirm, please explain fully.

(g) Did you consider and pursue any other methods for treating missing variable observations other than assigning zero values to the indicated variables? If so, please provide any results related to such analyses. If not, please explain why you did not pursue any alternative treatments.

(h) Please explain fully why in the indicated section of the SAS program you deleted observations with zero values for the street_hours, yet you assigned zero values to the volume variables when data for such variables were missing.

RESPONSE TO ADVO/OCA-T3-26.

(a) Please see my answer to USPS/OCA-T3-8 (b)-(c).

(b) No. Also, please see my answer to USPS/OCA-T3-8 (b)-(c).

RESPONSES OF OCA WITNESS J. EDWARD SMITH
TO INTERROGATORIES ADVO/OCA-T3-25-32

- (c) Yes.
- (d) On the assumption that the values *are* zero, my treatment is better. Also, please see my answer to USPS/OCA-T3-8 (b)-(c).
- (e) Confirmed.
- (f) Confirmed.
- (g) I have not yet done so as of the filing of this interrogatory. Since I received the DOIS data so close to the filing deadline for my testimony, time did not permit alternative treatments.
- (h) A regression with zero value for street_hours would be meaningless.

RESPONSES OF OCA WITNESS J. EDWARD SMITH
TO INTERROGATORIES ADVO/OCA-T3-25-32

ADVO/OCA-T3-27. Again, please refer to your SAS program ReadVolume contained in OCA-LR-L-4 and the section of your code prefaced by “the procedure below assumes that missing data is zero rather than deleting the data - this is an important assumption.” This specific assumption applies to missing data for the volume variables.

(a) Please explain fully why you consider this distinction and consequential treatment of the missing data an important assumption.

(c) Please explain if you expected any additional missing data (compared to those “cleaned up” earlier in the program) for the same variables to be present in the new Volume data set within this section and why.

(d) If you did not expect any new missing data to be present, why were values of zero reassigned once again to the same variables? Please explain fully.

(e) Please confirm that within the indicated section, two new variables, seq_letters_sets and seq_flats_pcs are now scanned for missing values and are assigned a value of zero for observations where missing data exist. If you cannot confirm, please explain.

(f) If you do confirm, please explain why these two variables were not included in the original scan referenced in the previous question?

(g) If there were actually missing non-zero values for those particular observations, please explain how you believe the models would be affected.

Response to ADVO/OCA-T3-27.

(a) The assumption will determine the inclusion/exclusion of a number of observations in the database.

(b) There is no (b) subpart.

(c) No. This is my programming preference.

(d) This is my programming preference.

(e) Confirmed.

(f) Neither of the variables is actually used in this program. They were simply included under a blanket inclusion.

RESPONSES OF OCA WITNESS J. EDWARD SMITH
TO INTERROGATORIES ADVO/OCA-T3-25-32

(g) Neither variable is used in the form listed. There would be no effect.

RESPONSES OF OCA WITNESS J. EDWARD SMITH
TO INTERROGATORIES ADVO/OCA-T3-25-32

ADVO/OCA-T3-28. Please refer again to your SAS program ReadVolume contained in Section 2 of OCA-LR-L-4. The last section of your program is "Final Data Cleaning".

(a) Please confirm that the purpose of this section of the program is:

1) to delete all observations containing either missing or zero valued data for the DELT variable, and 2) to reassign a zero value to all other variables identified in this section containing missing data. If you cannot confirm, please explain.

(b) Please confirm that, within this last section and with the exception of the DELT variable, the same variables identified in the previous two questions are again reassigned a zero value for observations in which their data are missing. If you cannot confirm, please explain.

(c) Please explain why you expected additional missing data for all variables indicated in this last section of the program to be present in the Volume data set.

(d) If you did not expect any new missing data to be present, why were values of zero reassigned once again to the same variables? Please explain fully.

(e) Please confirm that the variable seq_flats_sets, included in the SAS data file FNVOLADJ, filed as part of OCA LR L-4, is also included in the final VOLUME data set formed in the program, in the indicated last section of the program. If you cannot confirm, please explain.

(f) If you do confirm, please explain why the seq_flats_sets variable contained in this data set was not checked for missing data, as were the other indicated variables.

RESPONSE TO ADVO/OCA-T3-28.

(a) Confirmed.

(b) Confirmed

(c) I did not. This is my programming style.

(d) This is my programming style.

(e) Confirmed.

(f) The variable is at this point redundant.

RESPONSES OF OCA WITNESS J. EDWARD SMITH
TO INTERROGATORIES ADVO/OCA-T3-25-32

ADVO/OCA-T3-29. Please refer again to your SAS program ReadVolume contained in Section 2 of OCA LR L-4. After you merge your Volume data set with the CombinedZips data set containing a distribution of possible delivery values by type of delivery for each zip and route, you include the following lines of SAS code:

DATA volume;

SET volume;

if residential_curb = "." then delete;

(a) Please confirm that the new data set Volume referenced in the DATA statement, now excludes all zip-route-day observations where the residential-curb variable contains missing data. If you cannot confirm, please explain.

(b) Please explain why you did not delete any observations with missing values for the remaining seven possible delivery values: residential NDCBU, residential central, residential other, business curb, business NDCBU, business central and business other. Did you run a separate check on these variables not included in the SAS code? Please explain fully.

RESPONSE TO ADVO/OCA-T3-29.

(a) Confirmed

(b) An inspection of the data indicated that the earlier deletion would handle all cases.

RESPONSES OF OCA WITNESS J. EDWARD SMITH
TO INTERROGATORIES ADVO/OCA-T3-25-32

ADVO/OCA-T3-30. Please refer to the SAS data file fivoladj, filed as part of Section 2 in OCA LR L-4. Please explain the distinction between the SEQ_FLATS_PCS and SEQFLTS variables included in the file.

RESPONSE TO ADVO/OCA-T3-30

The variables are identical. SEQ_Flats_PCS has not been used as a variable.

RESPONSES OF OCA WITNESS J. EDWARD SMITH
TO INTERROGATORIES ADVO/OCA-T3-25-32

ADVO/OCA-T3-31. Please refer to the SAS data file fnlvoladj, filed as part of OCA LR L-4.

(a) Please explain your understanding of the MILES and LAND variables indicated in the file. Does the MILES variable include all route miles covered by the carrier to complete delivery (travel, run time, traveling between route sections, etc.) or just a portion of these miles? Does the LAND variable include all square miles for zip codes in which routes are located, or just a portion of these miles? Please explain fully.

(b) Do you believe that, among other factors, the sum of route miles for all routes within a particular zip code would be influenced by that zip's square miles in area? Please explain fully.

(c) Suppose volume attributable to a particular zip code suddenly doubled requiring an increase in routes, but no increase in the number of zip-codes serving the defined geographic area. Please explain fully what influence you would expect the increased workload to have on the MILES variable for that zip code.

RESPONSE TO ADVO/OCA-T3-31.

(a) Miles are base vehicle miles. See interrogatory OCA/USPS-T14-8. Land is square miles in the ZIP code and is exclusive of water.

(b) I don't know.

(c) I don't know what effect a doubling of volume to a particular ZIP code would have on numbers of routes or whether the ZIP code structure would be unaffected. The matter is purely speculative. Speculating, I tend to think that miles would not increase substantially, but I may be wrong.

RESPONSES OF OCA WITNESS J. EDWARD SMITH
TO INTERROGATORIES ADVO/OCA-T3-25-32

ADVO/OCA-T3-32. Please confirm that OCA LR L-4, Section 2, fnlvoladj SAS data set is the data set used to develop your DOIS models (described in your Section 3 SAS programs as newdois.fnlvolume) and that it is the same as the "volume" data set generated by the ReadVolume SAS program in the same section. If this is incorrect, please provide the correct data set in machine-readable format.

RESPONSE TO ADVO/OCA-T3-32.

Partly confirmed and partly denied. The results can be generated from either database;

I used newdois.fnlvolume. Unencoded ZIP code information has been deleted from fnlvoladj. Otherwise the databases are identical.