

Revising the Estimation Strategy for the Pesticide Data Program

Phillip S. Kott

The Pesticide Data Program (PDP) is a cooperative effort of the U.S. Department of Agriculture (USDA) and several state agencies. The ultimate purpose of the PDP is to make scientific statements about the distribution of certain pesticide residues in particular commodities consumed by the US public over a calendar year. The estimation strategy that has been used in the PDP for the last several years will be described and then the current (as of 2006) estimation routine will be compared to a new approach to be introduced in 2007.

KEY WORDS: Analysis weight; Benchmark; Level of detection; Level of quantification; Mean; Percentile; Population unit.

Phillip S. Kott is Chief Research Statistician, Research and Development Division, National Agricultural Statistics Service, 3251 Old Lee Highway, Room 305, Fairfax, Virginia, 22030.

1. Introduction

The Pesticide Data Program (PDP) is a cooperative effort of the U.S. Department of Agriculture (USDA) and several state agencies. The ultimate purpose of the PDP is to make scientific statements about the distribution of certain pesticide residues in particular commodities consumed by the US public over a calendar year. Usually these commodities are fresh or processed fruits or vegetables. The particular mix of commodities varies from year to year. The pesticides tested depend on capabilities of PDP-associated laboratories

The original theory governing sampling and estimation for the PDP is described in Kott and Carr (1997). Although the selection of states, sites, and samples has remained roughly the same as described there, the determination of national estimates has changed considerably over the years in a number of ways. These changes will be outlined in Section 2. The most important of these changes was abandoning the collecting of weekly commodity sales from wholesalers and relying instead on monthly “arrivals” data published by the Agricultural Marketing Service (AMS) of USDA. These data were used to adjust the samples of the fresh commodities to reflect more accurately what was distributed (and consumed) in the US in a year.

Section 3 described a new estimation methodology that will be implemented officially starting with the 2007 PDP data. This methodology uses place-of-origin information rather than month to adjust the sample. A set of national estimates has been calculated for 2006 with the new methodology. This allows a comparison of the results from using the new and old approaches.

The fresh commodities analyzed for pesticide residues in 2006 were bananas, carrots, cauliflower, cranberries, eggplant, grapefruit, plums, and winter squash. Also analyzed were orange juice, raisins, frozen potatoes, and frozen peas. Among processed commodities, only orange juice sample data were weighted. The goal in this being to reflect the three-to-one ready-to-serve over frozen ratio in the quantity of orange juice distributed in a year according to information provided by the Food Institute.

For each commodity, such as bananas, two distinctions need to be drawn. The first is between the *sample* of bananas analyzed by the PDP and the *target* population about which statistical statements can reasonably be made. The target population for the 2006 PDP was all bananas distributed to supermarkets, independent grocers, hotels, restaurants, institutions, and (on rare occasion) final consumers through wholesalers in 2006.

The second distinction is between this target population and the *inferential* population of the bananas actually consumed by the US public in 2006. This is not to say that there exists significant differences between the distributions or pesticide residues on the target and inferential populations. On the contrary, the USDA assumes that such differences are so minor that they can be ignored. Nevertheless, it is important to recognize that this assumption is being made. Like many assumption, it is a simplification. Useful, but potentially a cause of error.

Before leaving the question of the population of interest, it is important to understand that the population **unit** under discussion is *not* a single discrete element of the commodity in question (e.g., a particular banana). Rather, it is a homogenized “sample” of the product that varies in size by commodity (five pounds for bananas). Commodity sample sizes are in large part dictated by laboratory-testing considerations.

2. Computing the 2006 National Estimates

The official 2006 national estimates for select pesticide residues on particular commodities can be found in Appendix J of AMS (2007). A few population distributions are plotted in Appendix K as well. The table in Appendix J provides national estimates for selected commodity/pesticide pairs of

- the proportion of units with detectable residues,
- the average (mean) residue concentration in parts per million (expressed as a range as explained below), and
- the 50th, 75th, and 90th percentiles of the population residue distribution.

See Kott and Carr (1977) for a discussion of the impact on estimation of nonzero levels of pesticide detection (LODs) on the one hand and of levels of minimum quantification (LOQs) higher than these LODs on the other. The varying-LOD-across-labs problem addressed in Kott and Carr has been all but eliminated in recent years. At most two different labs were used to analyze a single commodity/pesticide pair in 2006.

The data on which the estimates in the table were derived were collected in nine states: New York, Maryland, Florida, Ohio, Michigan, Wisconsin, Texas, Colorado, Washington, and California. The assumption was that, for the most part, US wholesalers distribute (and consumers eat) bananas from Central America and carrots from California wherever they reside, so that the state where a wholesaler does business is less important than the month in which the product is sampled. Particular months are dominated by certain growing regions, and pesticide residuals largely reflect region of origin.

As in Kott and Carr, sample observations with residues between the LOD and LOQ were treated as four analysis observations with residue values equally spaced between the LOD and LOQ (literally at 35, 55, 75, and 95% of the LOQ, the LOD generally being 30% of the LOQ). All other sample observations were treated as four analysis observations with the same residue values.

Focusing on a particular fresh-commodity/pesticide pair, let N_j be, in principle, the total amount of the commodity distributed in month j of 2006, and n_j be the number of relevant PDP sample observations for that month.

In Kott and Carr, the **analysis weight** for a sample observation was set roughly proportional to the inverse of the probability of selecting it for the sample. This proved an extremely difficult value to determine in practice. As a result, the weight for an *analysis* observation in month j of 2006 was set instead at

$$w_k = \frac{1}{4} \frac{\left(\frac{N_j}{n_j} \right)}{\sum_{g=1}^{12} N_g}. \quad (1)$$

By applying equation (1), the fraction of the estimated population of the commodity distributed in month j becomes $N_j / \sum_{g=1}^{12} N_g$ – the actual fraction. The sample is said to be “benchmarked” to the 12 monthly quantities, N_j . (This weighting scheme has been used for computing national estimates of pesticide residues in fresh commodities since at least 1999.)

Observe that the analysis weight in equation (1) do not change when all the N_g are multiplied by a constant. Consequently, the commodity sizes measured by the N_g need only reflect relative differences across the months. To serve as a proxies for these relative measures, USDA used US-aggregated monthly *arrivals* data collected in 20 cities for 1998 (AMS, 1999). This was the last year for which such data were available.

Official nation estimates were computed as follows. By summing the w_k across all sampled units with positive residues, the USDA estimated the population proportion with positive residues.

Population means were estimated by computing w_k -weighted averages of the residue concentrations in the sample. The minimum value of the mean was computed by treating nondetects as if they had zero concentrations, while the maximum value of the mean treated nondetects as if they had concentrations equal to the LOD.

Weighted percentiles were estimated analogously to means using the FREQ statement in PROC UNIVARIATE of SAS® (SAS 2002). The use this statement effectively in this context, weights were first multiplied by 10,000.

3. A New Weighting Scheme

Extensive internal analyses of the 2005 PDP sample as well as the age of the *arrivals* data convinced USDA to revise the PDP weighting scheme. Instead of benchmarking to month, starting in 2007 USDA will benchmark to place-or-origin as best as can be determined. Moreover, the regions used as benchmarks (whether a country, a state, a group of countries, or a group of states) will vary from commodity to commodity depending on what data is available.

The new weighting formula is exactly the same as in equation (1), except now j (and g) indexes a region, one of G , where the size of G will depend on the commodity. For a sample in group j , the analysis weight will be

$$w_k = \frac{1}{4} \frac{\left(\frac{N_j}{n_j} \right)}{\sum_{g=1}^{12} N_g}. \quad (1)$$

where N_j is a measure of the amount of the commodity is question from region j .

Sometimes, as with orange juice in 2006, j will index groups defined using characteristics other than, or in addition to, region. In the case of orange juice, domestic versus imported origin was supplemented by whether the sample was ready-to-serve or frozen.

The results for 2006 commodity/pesticide pairs that were re-estimated using the new weighting scheme are displayed in Table 1. Estimates using both the old and new methods are reported.

As noted in Section 2, OLD estimates for fresh commodities were benchmarked to AMS monthly “arrivals” data, last collected in 1998. Among these, spinach estimates were based on sample data and benchmarks from January through October, while winter squash estimates were based on sample data and all-squash benchmarks from January through September. There were no separate *arrivals* numbers for winter squash.

NEW estimates were benchmarked to place-of-origin data, usually imported versus domestically-originated product. The Economic Research Service (ERS) of USDA supplied 2006 import fractions for domestically-distributed commodities.

Some commodities were further benchmarked by state of origin using National Agricultural Statistics Service (NASS) data, with several states often grouped together (e.g., Florida and Georgia, or all states other than California). Similarly, some commodities were further benchmarked by country of origin using Foreign Agriculture Service (FAS) data. When this happened for a commodity it was usually either Canada versus the rest of the world or Mexico versus the rest of the world. Bananas being 100% imported were benchmarked to five specific countries of origin.

Most domestic data were for 2006, but 2002 Census-of-Agriculture data were used for eggplant because NASS does not collect annual eggplant numbers.

Over 5% of cranberry samples were marked imported on the PDP data set despite ERS estimates that no appreciable amount of fresh cranberries were imported in 2006. To decrease the impact of these samples but not ignore them entirely, we used 2% as the imported benchmark for cranberries. Other cranberry benchmarks were Massachusetts (44.8%) and all other states grouped together (53.2%). The lack of samples originating in Wisconsin (the leading state cranberry producer) suggests cranberry estimates should to be used with caution.

Recall that in the OLD approach, orange juice was benchmarked using two categories: ready-to-serve and frozen. The NEW approach used four categories: ready-to-serve/domestic, frozen/domestic, ready-to-serve/imported, frozen/imported. Data for this came from both ERS (on 2006 imports) and NASS (on the utilization of the 2005/2006 and 2006/2007 domestic crops).

Winter squash was benchmarked to all-squash (both winter and summer varieties) numbers. This is the only level of aggregation at which NASS and ERS collect and publish data.

Slightly under 5% of domestic samples and 1.5% of imported samples had no state/country of origin. To use these records, samples marked as domestic (imported) but without a specific originating state (country) were prorated across the benchmarking states (countries) using averages within the collection state and season. For example,

say 40% of the winter sample of a commodity collected in Texas and having a marked state of origin came from California (and 60% came from other states), then a winter sample of that commodity collected in Texas and marked as domestic but without a state of origin was treated as 40% from California (and 60% from other states). A similar process was used for samples with unknown import/domestic status (0.5% of the samples).

More information on the estimation process can be gleaned from the SAS code reproduced in an appendix appearing after Table 1.

AMS (1999). *Fresh Fruit and Vegetable Arrival Totals for 20 Cities: FVAS-3 Calendar Year 1998*. Fruit and Vegetable Programs: Market News Branch. Agricultural Marketing Service, USDA.

AMS (2007). *Pesticide Data Program: Annual Summary, Calendar Year 2006*. Science and Technology Programs. Agricultural Marketing Service, USDA.

Kott, P. S. and Carr, D. A. (1997). "Developing an Estimation Strategy for a Pesticide Data Program." *Journal of Official Statistics*, 13, 367-383.

SAS (2002). SAS 9. SAS Institute, Cary NC. <http://support.sas.com/documentation/>

**Table 1. Comparing Old and New National Estimation Methods for the 2006 PDP :
Pairs with Residue Detections in at Least 10 Percent of Samples¹**
(Means and percentiles in parts per million)

Commodity	Pesticide	Percent of Population with Detections	Mean**		Percentiles			Ratio of 90 th Percentile to Tolerance
			lower	upper	50 th	75 th	90 th	
2. Bananas								
OLD	Imazalil	31.0	0.006	0.012	*	0.010	0.019	0.006
	Thiabendazole	65.6	0.026	0.030	0.014	0.036	0.066	0.022
NEW	Imazalil	30.8	0.006	0.012	*	0.010	0.019	0.006
	Thiabendazole	65.4	0.026	0.030	0.014	0.036	0.066	0.022
3. Carrots								
OLD	Azoxystrobin	15.7	0.001	0.001	*	*	0.002	0.005
	DDE p,p'	27.1	0.003	0.005	*	0.003	0.011	0.004
	Iprodione	13.1	0.007	0.033	*	*	0.038	0.008
	Pyraclostrobin	37.5	0.002	0.003	*	0.002	0.007	0.018
	Trifluralin	57.4	0.021	0.025	0.003	0.034	0.066	0.066
NEW	Azoxystrobin	15.8	0.001	0.001	*	*	0.002	0.005
	DDE p,p'	27.5	0.003	0.005	*	0.003	0.011	0.004
	Iprodione	13.4	0.007	0.033	*	*	0.038	0.008
	Pyraclostrobin	37.6	0.002	0.003	*	0.003	0.007	0.018
	Trifluralin	57.7	0.021	0.025	*	0.034	0.068	0.068
4. Cauliflower								
OLD	Imidacloprid	76.4	0.002	0.002	0.001	0.003	0.007	0.002
NEW	Imidacloprid	73.7	0.002	0.002	0.001	0.003	0.007	0.002
5. Cranberries								
OLD	Azoxystrobin	31.4	0.001	0.001	*	0.001	0.004	0.008
	Chlorpyrifos	22.4	0.006	0.010	*	*	0.026	0.026
	Methoxyfenozide	14.4	0.002	0.007	*	*	0.009	0.018
NEW	Azoxystrobin	19.0	0.001	0.001	*	*	0.002	0.004
	Chlorpyrifos	14.9	0.004	0.008	*	*	0.011	0.011
	Methoxyfenozide	23.6	0.004	0.008	*	*	0.014	0.028

The numbers to the right of the commodities refer to their enumeration in Table J of AMS (2007).

See the final page of this table for explanations of the footnotes. Descriptions of the OLD and NEW weighting processes are in the text.

Commodity	Pesticide	Percent of Population with Detections	Mean **		Percentiles			Ratio of 90 th Percentile to Tolerance
			lower	upper	50 th	75 th	90 th	
6. Eggplant (W)								
OLD	Endosulfan II	13.8	0.002	0.006	*	*	0.008	0.008
	Endosulfan sulfate	17.3	0.004	0.009	*	*	0.015	0.015
NEW	Endosulfan II	14.0	0.002	0.006	*	*	0.008	0.008
	Endosulfan sulfate	17.6	0.004	0.009	*	*	0.015	0.015
7. Grapefruit								
OLD	Imazalil	38.4	0.022	0.046	*	0.029	0.067	0.007
NEW	Imazalil	36.3	0.023	0.049	*	0.025	0.070	0.007
8. Orange Juice								
OLD	Carbaryl	22.3	0.001	0.005	*	*	0.003	< 0.001
	Imazalil	11.3	0.003	0.024	*	*	0.006	0.001
NEW	Carbaryl	27.5	0.001	0.005	*	0.001	0.004	< 0.001
	Imazalil	10.4	0.002	0.024	*	*	0.004	< 0.001
10. Plums								
OLD	Chlorpyrifos	8.9	0.001	0.004	*	*	*	*
	Fludioxonil	27.8	0.107	0.186	*	*	0.342	0.068
	Iprodione	10.9	0.107	0.142	*	*	0.098	0.005
	Phosmet	23.9	0.004	0.006	*	*	0.011	0.002
NEW	Chlorpyrifos	12.2	0.002	0.004	*	*	0.006	0.110
	Fludioxonil	23.9	0.093	0.177	*	*	0.342	0.068
	Iprodione	22.5	0.212	0.243	*	*	0.700	0.035
	Phosmet	18.8	0.004	0.006	*	*	0.010	0.002
11. Potatoes, frozen								
OLD	Chlorpropham	61.4	0.423	0.426	0.210	0.715	1.200	0.024
	Imidacloprid	33.1	0.002	0.005	*	0.003	0.007	0.018
NEW	Chlorpropham	62.3	0.438	0.442	0.220	0.730	1.200	0.024
	Imidacloprid	37.8	0.002	0.005	*	0.004	0.007	0.018

Commodity	Pesticide	Percent of Population with Detections	Mean**		Percentiles			Ratio of 90 th Percentile to Tolerance
			lower	upper	50 th	75 th	90 th	
12. Raisins								
OLD	Boscalid	35.8	0.017	0.034	*	0.015	0.055	0.006
	Propargite	31.5	0.030	0.059	*	0.020	0.089	0.009
	Trifloxystrobin	25.8	0.005	0.025	*	0.001	0.012	0.002
NEW	Boscalid	28.9	0.014	0.032	*	0.009	0.045	0.005
	Propargite	25.8	0.025	0.056	*	0.006	0.077	0.008
	Trifloxystrobin	25.0	0.005	0.025	*	0*	0.010	0.002
13. Spinach								
OLD	Azoxystrobin	10.6	0.041	0.042	*	*	0.001	< 0.001
	Cypermethrin	14.7	0.056	0.173	*	*	0.210	0.021
	DDE p,p'	37.6	0.003	0.005	*	*	0.010	0.020
	Imidacloprid	71.7	0.020	0.021	0.003	0.017	0.051	0.015
	Omethoate	13.1	0.014	0.017	*	*	0.006	0.003
	Permethrin cis	56.5	0.415	0.418	0.019	0.690	1.400	0.070
	Permethrin trans	56.2	0.413	0.415	0.019	0.630	1.300	0.065
NEW	Azoxystrobin	11.1	0.044	0.045	*	*	0.001	< 0.000
	Cypermethrin	15.1	0.057	0.170	*	*	0.200	0.020
	DDE p,p'	37.0	0.003	0.005	*	0.005	0.010	0.020
	Imidacloprid	75.1	0.025	0.026	0.005	0.024	0.060	0.017
	Omethoate	13.5	0.010	0.013	*	*	0.008	0.004
	Permethrin cis	60.4	0.457	0.459	0.055	0.740	1.400	0.070
	Permethrin trans	60.2	0.444	0.446	0.050	0.680	1.300	0.065
14. Sweet peas, frozen								
OLD	Dimethoate	12.1	0.001	0.005	*	*	0.005	0.002
NEW	Dimethoate	12.0	0.001	0.005	*	*	0.005	0.002
16. Winter squash								
OLD	Endosulfan sulfate	29.1	0.006	0.011	*	0.008	0.023	0.023
NEW	Endosulfan sulfate	29.8	0.007	0.012	*	0.008	0.025	0.025

The footnotes

1. Includes some pairs with detections in 10 percent of the samples but with estimated detections in less than 10 percent of the population. Excludes pairs with less than 100 sample observations, without observations in all participating states, or with observations from less than six months of the year (except for fresh cranberries, which are available almost exclusively in the winter).

* The percentile values is estimated either to be below the level of detection (LOD).

** The mean is estimated with a range of values. The lower bound is calculated with non-detects valued at zero. The upper bound is calculated using the LOD.

NT No tolerance.

APPENDIX: SAS Code for New Weighting Method with 2006 PDP Sample

```
* Handles missing origin codes;

IF ORIGSAT = '' THEN ORIGSAT = 'UN'; * ORIGSAT used in place of ORIGSTAT;
IF COUNTRY = '' THEN COUNTRY = 'UNK';

* Creates seasons

IF MO < 4 THEN SEASON = 1; ELSE IF MO < 7 THEN SEASON = 2; ELSE IF MO < 10 THEN SEASON = 3;
ELSE SEASON = 4;

* Creates analysis sets and imputation sets;

DATA S I MS MI M; SET D; * S - Domestic I - Imports MS - Domestic/missing-state
                        MI - Import/missing-country M - Missing import status;

GS1 = 0; GS2 = 0 ; GS3 = 0 ; GSO = 0; * 0 MEANS OTHER;
GI1 = 0; GI2 = 0 ; GI3 = 0 ; GI4 = 0; GI5 = 0; GIO = 0; * 0 MEANS OTHER;

IF ORIGIN = 1 AND ORIGSAT = 'UN' THEN OUTPUT MS; * MISSING DOMESTIC ;
IF ORIGIN = 2 AND COUNTRY = 'UNK' THEN OUTPUT MI; * MISSING IMPORT ;
IF ORIGIN = 3 THEN OUTPUT M; * IMPORT STATUS UNKNOWN;

IF PR = 'BN' THEN DO; * Bananas;
  IF COUNTRY = '285' THEN DO; GI1 = 1; OUTPUT I; END; * Columbia;
  IF COUNTRY = '295' THEN DO; GI2 = 1; OUTPUT I; END; * Costa Rica;
  IF COUNTRY = '325' THEN DO; GI3 = 1; OUTPUT I; END; * Ecuador;
  IF COUNTRY = '415' THEN DO; GI4 = 1; OUTPUT I; END; * Guatemala;
  IF COUNTRY = '430' THEN DO; GI5 = 1; OUTPUT I; END; * Honduras;
  IF GI1 + GI2 + GI3 + GI4 + GI5 = 0 AND COUNTRY NE 'UNK' THEN DO; GIO = 1; OUTPUT I; END;
  * Other Import;
END;
```

```

IF PR = 'CR' THEN DO; * Carrots;
  IF ORIGSAT = 'CA' THEN DO; GS1 = 1; OUTPUT S; END; * CA;
  IF ORIGSAT = 'MI' THEN DO; GS2 = 1; OUTPUT S; END; * MI;
  IF ORIGSAT = 'TX' THEN DO; GS3 = 1; OUTPUT S; END; * TX;
  IF ORIGIN = 1 AND GS1 + GS2 + GS3 = 0 AND ORIGSAT NE 'UN' THEN DO; GSO = 1; OUTPUT S; END;
  * OTHER DOMESTIC;
  IF ORIGIN = 2 AND COUNTRY = '260' THEN DO; GI1 = 1; OUTPUT I; END; * CANADA ;
  IF ORIGIN = 2 AND COUNTRY NE '260' AND COUNTRY NE 'UNK' THEN DO; GI2 = 1; OUTPUT I; END; * OTHER IMPORT;
END;

IF PR = 'CA' THEN DO; * Cranberries
  IF ORIGIN = 1 AND ORIGSAT = 'MA' THEN DO; GS1 = 1; OUTPUT S; END; * MA;
  IF ORIGIN = 1 AND ORIGSAT NE 'MA' AND ORIGSAT NE 'UN' THEN DO; GS2 = 1; OUTPUT S; END; * OTHER;
  IF ORIGIN = 2 THEN DO; GI1 = 1; OUTPUT I; END; * IMPORT;
END;

IF PR = 'CF' THEN DO; * Cauliflower;
  IF ORIGIN = 1 AND ORIGSAT = 'CA' THEN DO; GS1 = 1; OUTPUT S; END; * CALIFORNIA;
  IF ORIGIN = 1 AND ORIGSAT NE 'CA' AND ORIGSAT NE 'UN' THEN DO; GSO = 1; OUTPUT S; END; * OTHER DOMESTIC;
  IF ORIGIN = 2 THEN DO; GI1 = 1; OUTPUT I; END; * IMPORT;
END;

IF PR = 'EP' THEN DO; * Eggplant;
  IF ORIGIN = 1 AND ORIGSAT = 'CA' THEN DO; GS1 = 1; OUTPUT S; END; * CALIFORNIA;
  IF ORIGIN = 1 AND (ORIGSAT = 'GA' OR ORIGSAT = 'FL') THEN DO; GS2 = 1; OUTPUT S; END; * GA/FL;
  IF ORIGIN = 1 AND GS1 + GS2 = 0 AND ORIGSAT NE 'UN' THEN DO; GSO = 1; OUTPUT S; END;
  * OTHER DOMESTIC;
  IF ORIGIN = 2 AND COUNTRY = '595' THEN DO; GI1 = 1; OUTPUT I; END; * MEXICO;
  IF ORIGIN = 2 AND COUNTRY NE '595' AND COUNTRY NE 'UNK' THEN DO; GIO = 1; OUTPUT I; END;
  * OTHER INPORT;
END;

IF PR = 'GF' THEN DO; * Grapefruit;
  IF ORIGSAT = 'CA' THEN DO; GS1 = 1; OUTPUT S; END; * CA;
  IF ORIGSAT = 'FL' THEN DO; GS2 = 1; OUTPUT S; END; * FL;
  IF ORIGSAT = 'TX' THEN DO; GS3 = 1; OUTPUT S; END; * TX;
  IF ORIGIN = 1 AND GS1 + GS2 + GS3 = 0 AND ORIGSAT NE 'UN' THEN DO; GSO = 1; OUTPUT S; END;
  * OTHER DOMESTIC;
  IF ORIGIN = 2 THEN DO; GI1 = 1; OUTPUT I; END; * IMPORT;
END;

IF PR = 'PU' THEN DO; * Plums;

```

```

    IF ORIGIN = 1 THEN DO; GS1 = 1; OUTPUT S; END; * DOMESTIC;
    IF ORIGIN = 2 THEN DO; GI1 = 1; OUTPUT I; END; * IMPORT;
END;

IF PR = 'PZ' THEN DO; * Froxen potatoes;
    IF ORIGIN = 1 THEN DO; GS1 = 1; OUTPUT S; END; * DOMESTIC;
    IF ORIGIN = 2 THEN DO; GI1 = 1; OUTPUT I; END; * IMPORT;
END;

IF PR = 'RA' THEN DO; * Raisins;
    IF ORIGIN = 1 THEN DO; GS1 = 1; OUTPUT S; END; * DOMESTIC;
    IF ORIGIN = 2 THEN DO; GI1 = 1; OUTPUT I; END; * IMPORT;
END;

IF PR = 'SP' THEN DO; * Spinach
    IF ORIGIN = 1 AND ORIGSAT = 'CA' THEN DO; GS1 = 1; OUTPUT S; END; * CALIFORNIA;
    IF ORIGIN = 1 AND ORIGSAT NE 'CA' AND ORIGSAT NE 'UN' THEN DO; GSO = 1; OUTPUT S; END; * OTHER DOMESTIC;
    IF ORIGIN = 2 THEN DO; GI1 = 1; OUTPUT I; END; * IMPORTS;
END;

IF PR = 'PS' THEN DO; * Frozen peas;
    IF ORIGIN = 1 THEN DO; GS1 = 1; OUTPUT S; END; * DOMESTIC;
    IF ORIGIN = 2 THEN DO; GI1 = 1; OUTPUT I; END; * IMPORT;
END;

IF PR = 'WS' OR PR = 'SS' THEN DO; * Squash (no separate benchmark for winter squash);
    IF ORIGIN = 1 AND ORIGSAT = 'CA' THEN DO; GS1 = 1; OUTPUT S; END; * CALIFORNIA;
    IF ORIGIN = 1 AND (ORIGSAT = 'GA' OR ORIGSAT = 'FL') THEN DO; GS2 = 1; OUTPUT S; END; * GA/FL;
    IF ORIGIN = 1 AND GS1 + GS2 = 0 AND ORIGSAT NE 'UN' THEN DO; GSO = 1; OUTPUT S; END; * OTHER DOMESTIC;
    IF ORIGIN = 2 THEN DO; GI1 = 1; OUTPUT I; END; * IMPORT;
END;

```

```

IF PR = 'OJ' THEN DO; * Orange juice;
  IF SEASON = 1 AND ORIGIN = 1 THEN DO; GS1 = 1; OUTPUT S; END; * FROZEN DOMESTIC;
  IF SEASON NE 1 AND ORIGIN = 1 THEN DO; GS2 = 1; OUTPUT S; END; * Ready-to-serve DOMESTIC;
  IF SEASON = 1 AND ORIGIN = 2 THEN DO; GI1 = 1; OUTPUT I; END; * FROZEN IMPORTED;
  IF SEASON NE 1 AND ORIGIN = 2 THEN DO; GI2 = 1; OUTPUT I; END; * Ready-to-Serve IMPORTED;
END;

```

* Organizes the sets for handling imputations;

```

proc sort DATA = MS; by pr pes_code STATE SEASON;
proc sort DATA = MI; by pr pes_code STATE SEASON;
proc sort DATA = S; by pr pes_code STATE SEASON;
proc sort DATA = I; by pr pes_code STATE SEASON;

```

```
DATA MI; SET MI; DROP GI1 GI2 GI3 GI4 GI5 GIO;
```

```
DATA MS; SET MS; DROP GS1 GS2 GS3 GSO;
```

```
DATA M; SET M; DROP GI1 GI2 GI3 GI4 GI5 GIO GS1 GS2 GS3 GSO;
```

* Takes the mean in the INPUT set by state and season;

```

PROC MEANS DATA = I NOPRINT; BY PR PES_CODE STATE SEASON; VAR GI1 GI2 GI3 GI4 GI5 GIO;
  OUTPUT OUT = OUTI MEAN = GI1 GI2 GI3 GI4 GI5 GIO;

```

* Takes the mean in the DOMESTIC set by state and season;

```

PROC MEANS DATA = S NOPRINT; BY PR PES_CODE STATE SEASON; VAR GS1 GS2 GS3 GSO;
  OUTPUT OUT = OUTS MEAN = GS1 GS2 GS3 GSO;

```

* Applies the state/season mean in the IMPORT (DOMESTIC) set to the missing set;

```

DATA MI; MERGE MI OUTI; BY PR PES_CODE STATE SEASON; IF ORIGIN = . THEN DELETE;
DATA MS; MERGE MS OUTS; BY PR PES_CODE STATE SEASON; IF ORIGIN = . THEN DELETE;

```

* Merges sets as of now in anticipation of dealing with M;

```

DATA D; SET S I MS MI;

PROC SORT DATA = D; by pr pes_code STATE MO;

* Takes tne State/Season mean;

PROC MEANS NOPRINT; BY PR PES_CODE STATE SEASON; VAR GI1 GI2 GI3 GI4 GI5 GIO GS1 GS2 GS3 GSO;
  OUTPUT OUT=OUT MEAN =GI1 GI2 GI3 GI4 GI5 GIO GS1 GS2 GS3 GSO;

* Applies mean to M;

DATA M; MERGE M OUT; BY PR PES_CODE STATE SEASON; IF P = . THEN DELETE;

* Merges

DATA D ; SET D M;

PROC SORT DATA = D; BY PR PES_CODE STATE MO ORIGIN;

* Computes means and counts. LOD/LOQ observations created after weighting;

PROC MEANS NOPRINT; BY PR PES_CODE; VAR GI1 GI2 GI3 GI4 GI5 GIO GS1 GS2 GS3 GSO;
  OUTPUT OUT = COUNT SUM = CI1 CI2 CI3 CI4 CI5 CIO CS1 CS2 CS3 CSO;

```


* Creates the weights.

Example: 143274 is the N_1 for banana imported from Columbia.

The weights must be big, but not too big, for SAS;

```
IF PR = 'BN' THEN DO;
```

```
WGT = (143274 * GI1/CI1) + (272136 * GI2/CI2) + (289110 * GI3/CI3) + (237854 * GI4/CI4) +  
(116390 + GI5/CI5) + (424180 *GIO/CIO);
```

```
WGT = WGT /1000;
```

```
END;
```

```
IF PR = 'CR' THEN DO;
```

```
WGT = (44757 * GI1/CI1) + (11826 * GI2/CI2) + (444050 * GS1/CS1) + (14400 * GS2/CS2) + (12338 + GS3/CS3) +  
(68240 *GSO/CSO);
```

```
WGT = WGT/1000;
```

```
END;
```

```
IF PR = 'CA' THEN DO;
```

```
WGT = (44800 * GS1/CS1) + (53200 * GS2/CS2) + (02000 * GI1/CI1);
```

```
END;
```

```
IF PR = 'CF' THEN DO;
```

```
WGT = (9276 * GI1/CI1) + (205482 * GS1/CS1) + (42919 *GSO/CSO);
```

```
WGT = WGT ;
```

```
END;
```

```
IF PR = 'EP' THEN DO;
```

```
WGT = (4277 * GI1/CI1) + (814 * GIO/CIO) + (1475 * GS1/CS1) + (2294 * GS2/CS2) + (2632 * GSO/CSO);
```

```
WGT = WGT * 10 ;
```

```
END;
```

```
IF PR = 'GF' THEN DO;
```

```
WGT = (903 * GI1/CI1) + (5300 * GS1/CS1) + (6914 * GS2/CS2) + (3200 * GS3/CS3) + (100 * GSO/CSO);
```

```
WGT = WGT * 10;
```

```
END;
```

```
IF PR = 'PU' THEN DO;
```

```
WGT = (77500 * GS1/CS1) + (22500 * GI1/CI1);
```

```
END;
```

```
IF PR = 'PS' THEN DO;  
WGT = (88700 * GS1/CS1) + (11300 * GI1/CI1);  
END;
```

```
IF PR = 'PZ' THEN DO;  
WGT = (78200 * GS1/CS1) + (21800 * GI1/CI1);  
END;
```

```
IF PR = 'SP' THEN DO;  
WGT = (1473 * GI1/CI1) + (138600 * GS1/CS1) + (43165 * GSO/CSO);  
WGT = WGT;  
END;
```

```
IF PR = 'RA' THEN DO;  
WGT = (86000 * GS1/CS1) + (14000 * GI1/CI1);  
END;
```

```
IF PR = 'OJ' THEN DO;  
WGT = (22000 * GI1/CI1) + (3000 * GI2/CI2) + (21000 * GS1/CS1) + (54000 * GS2/CS2);  
END;
```

```
IF PR = 'WS' OR PR = 'SS' THEN DO;  
WGT = (144207 * GI1/CI1) + (37929 * GS1/CS1) + (88680 * GS2/CS2) + (102777 * GSO/CSO);  
WGT = WGT / 10;  
END;
```