

Data Management – QAQC

Claire Osgood
November 2017



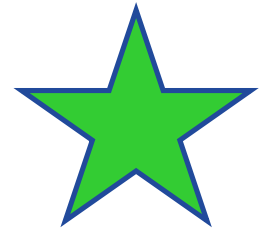
QAQC and Programming Standards



QAQC and Programming Standards



All work is verified/checked by a second person



- Like proofing a report/paper
- Use Word with Track Changes to check programs
- Manual data entry has a different process
- Includes checking documentation
- “Freeze” when fully verified (program, logs, listings/prints, input and output files)

QAQC and Programming Standards – Code Checking

```

set ds0106_w_trt;
STATEFP00=put(substr(CTIDFP00,1,2),2.);
TRACTCE00=put(substr(CTIDFP00,6,6),6.);

run;

*Running a check. For example, Census 2000 tract 37183054201 contains 6 07/08 downscaler points. Is this in the
data?;
proc freq data=ds0708_w_trt (where=(CTIDFP00="37183054201"));
title2 "Making sure this tract appears 6 times in the dataset";
tables CTIDFP00;

run;






**
** Macro:
** For each time period, combine the tracts that contain a DS point with those that
** are assigned the nearest DS point.
** -- Flag the source dataset
** -- Convert distance from feet to miles
** -- Do some checks
**
** Input parameters: DS dataset name, Tract dataset name, Output dataset name.
**
%macro set (ds, trt, out);
data &out;
set &ds (in=ds_ptsx) &trt (in=centrtx);
if TRACTCE00^='000000' then coast=0;
else coast=1;
ds_pts=ds_ptsx;
centrt=centrtx;
if NEAR_DIST=. then NEAR_DIST=0;
near_dist_miles=NEAR_DIST*0.000189394;

run;

proc freq data=&out;
title3 "Number of 2000 Census tracts on coast (coast=1)";
tables coast/list missing; *8;

run;

```

- 
Claire E Osgood
 I would like to see how many tracts are associated with more than one DS point. Please add that check.
- 
Julie Strominger
 Resolved.
- 
Julie Strominger
 Resolved.
- Claire E Osgood**
 Deleted: set together so know which ds points contribute to each Census tracts, ¶
- Claire E Osgood**
 Moved (insertion) [1] →
- Claire E Osgood**
 Deleted: →
- Claire E Osgood**
 Deleted: *Macro to set DS and Tract dataset together, flag which dataset the observation came from, ¶
- Claire E Osgood**
 Moved up [1]: → Input: DS dataset name, Tract dataset name, Output dataset name.; ¶ →
- 
Claire E Osgood
 By definition, these are exact opposite in value. Seems redundant. Do you really want to know the ¶
- 
Julie Strominger
 Resolved. Created one indicator variable (contain_ds_pts).

QAQC and Programming Standards

Checking Programs

- Comments and Titles
- Data Steps and Merges
- Data Manipulation
- Audit Trail

Resource: [Program Checks.docx](#)



QAQC and Programming Standards – Checking Programs

Data Steps and Merges

- # of Observations going into and out of data steps
- # of Variables going into and out of data steps
- LOG checks – key words ERROR, WARNING, INVALID, MISSING, etc.)
- Merges

UniqueID	Race
1	1
2	1
3	2
4	5
5	3

UniqueID	Race
1	A
2	
3	B
4	C
5	H

Merge will overwrite Race from one dataset with Race from the other.



UniqueID	Week
1	1
1	2
1	3
2	1
2	2

UniqueID	Address
1	Addr 1
1	Addr 2
2	Addr 1
3	Addr 1
4	Addr 1

Many-to-many merge. Is this what you want? What will the output look like?



QAQC and Programming Standards – Checking Programs

Data Manipulation – Use Freqs to show variable value conversions


Race	Race_Code	Frequency	Percent	Cumulative Frequency	Cumulative Percent
		1213976	44.32	1213976	44.32
	?	4	0.00	1213980	44.32
	[1	0.00	1213981	44.32
	_	1	0.00	1213982	44.32
0	0	11	0.00	1213993	44.32
1	1	14398	0.53	1228391	44.85
2	2	29706	1.08	1258097	45.93
3	3	557463	20.35	1815560	66.29
4	4	922	0.03	1816482	66.32
5	5	857955	31.33	2674437	97.65
6	6	41	0.00	2674478	97.65
7	7	14587	0.53	2689065	98.18
8	8	14561	0.53	2703626	98.71
9	9	12842	0.47	2716468	99.18
A	A	162	0.01	2716630	99.19
A	a	4	0.00	2716634	99.19
B	B	1244	0.05	2717878	99.23
B	b	95	0.00	2717973	99.24
C	C	3860	0.14	2721833	99.38
C	c	304	0.01	2722137	99.39

QAQC and Programming Standards – Checking Programs

Data Manipulation – Use Prints to show variable value calculations

Age Calculations, as of 1/1/2016

<u>BirthDate</u>	Age
.	.
.	.
.	.
.	.
.	.



Age Calculations, as of 1/1/2016

<u>BirthDate</u>	Age
08/08/1964	51
12/07/2000	15
05/05/1999	16
03/28/1995	20
09/24/1985	30

