

Errata for Fundamentals of Programming in SAS: A Case Studies Approach

1. Chapter 1

- (a) No errors noted.

2. Chapter 2

- (a) All Section 2.2 output is built from observations for which mortgage payment is \$100 per month or more, not just Output 2.2.1.
- (b) Program 2.7.7 is not correct in the book, the correct program (which is in the code that accompanies the book) is:

```
proc freq data=bookdata.ipums2005basic;
  table HomeValue*Metro/ nocol nopercnt format=dollar14.;
  weight MortgagePayment;
  format Metro MetroB. HomeValue Hvalue. ;
run;
```

- (c) Program/Log 2.9.4 do not match - the PUTLOG statement that displays the values for FlightNum and Date has the prefix NOTE: in the code, but not in the log.

3. Chapter 3

- (a) After Output 3.8.1 (page 124), FIRSTOBS= is referred to as a "familiar option"; however, it has not been seen before this point in the book.
- (b) In the raw file IPUMS2005Dirtied.dat, the serial and state postal code are not delimited from each other, so Program 3.9.1 reads them together in the column named Serial in the INPUT statement. This is not displayed in any of the subsequent work, but may cause confusion. It was intended that splitting these into two proper columns be left as an exercise, but it was not included.
- (c) In Programs 3.9.9 and 3.9.10, DOLLAR20. is used as the informat for the INPUT function; however, it is technically not an informat. Rather it is an alias to COMMA20., which is preferable usage to match the informats used for HHIncome and HomeValue in the modified list input in those same programs.

4. Chapter 4

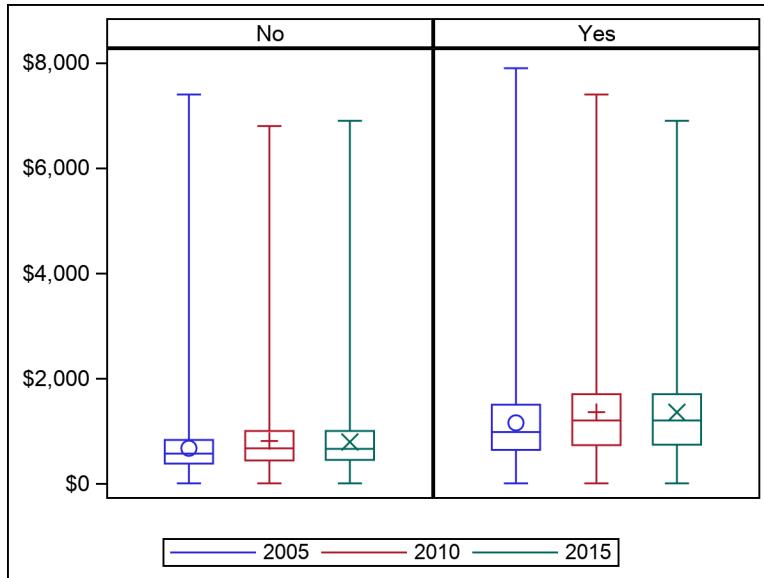


Figure 1: Output 4.2.4

- (a) Output 4.2.4 is not correct (the Yes/No flag on metropolitan status is not constructed correctly). The correct figure is shown above:
- (b) Program 4.4.3 Callout 4 is somewhat self-contradictory, here is a better description: By default, when applied to specific output objects, ODS SELECT and EXCLUDE are only in effect until the step boundary that ends the next procedure. Thus, the effect is local to the current (or next) procedure. Use of the keywords ALL or NONE to select output objects with ODS SELECT or EXCLUDE makes the effect global.
- (c) The RENAME= option applied to work.basic2010B in Program 4.4.8 should be:

```
rename= (MortStat=MortgageStatus
         MortPay=MortgagePayment
         Inc=HHIncome)
```

- (d) The assignment of MetFlag is not consistent with the data in Program 4.5.2 (nor with the formats constructed for it previously), a correct form of conditioning is:

```
if Metro in (2,3,4) then MetFlag = 'Yes';
else if Metro eq 1 then MetFlag = 'No';
else MetFlag = 'N/A';
```

- (e) For Callout 5 in Program 4.5.3, it is helpful to refer back to the properties for WHERE conditioning listed at the start of Section 2.6 (page 54).

5. Chapter 5

- (a) The WHERE condition in Program 5.3.5 does not match the output (the code given with the book is correct, however). The correct WHERE condition is:

```
where state in ('North Carolina', 'South Carolina')  
    and mortgageStatus ne 'N/A';
```

- (b) The data set IPUMS2005Cost is generated in Program 5.3.5, but not actually used there. It is first used in Program 5.3.6.
- (c) In short answer question 5a, the phrase "user-defined variables" may cause a bit of confusion, "data set variables" is what is intended.
- (d) In short answer question 7, the final statement prior to the RUN statement should be:

```
if Age eq 30 then output CompanyB;
```

6. Chapter 6

- (a) Programs 6.3.3, 6.3.4, and 6.3.5 use a DATA step SUM statement in the DO loop, but tit is not introduced until Chapter 7.
- (b) Program 6.3.3 uses the variable names North_Carolina and South_Carolina in reference to the transposed data set, Means2. This is correct if VALIDVARNAME=V7 is active. If VALIDVARNAME=ANY, as is the default in SAS Studio, the naming needs to change in the DATA step and PROC SGLOT ('North Carolina'n and 'South Carolina'n are the names produced by PROC TRANSPOSE in this case). There is no chapter note about this until Chapter 7, and it does not provide extensive detail.
- (c) The first ARRAY statement in short answer question 5a is missing a semicolon at the end.

7. Chapter 7

- (a) Output 7.2.4 has the overall mean of all data from both 2005 and 2010 in the column labeled 2005 – this column can be omitted altogether, or it can be replaced with the actual 2010 mean, which is the basis for the color coding. The table is:

		Projected Mean Cost				
State	Actual 2010 Mean	2006	2007	2008	2009	2010
Connecticut	\$1,746.45	\$1,451.46	\$1,480.49	\$1,510.10	\$1,540.30	\$1,571.11
Massachusetts	\$1,242.80	\$1,196.78	\$1,220.71	\$1,245.13	\$1,270.03	\$1,295.43
New Jersey	\$1,888.13	\$1,507.42	\$1,537.57	\$1,568.32	\$1,599.69	\$1,631.68
New York	\$1,528.60	\$1,483.44	\$1,513.11	\$1,543.37	\$1,574.24	\$1,605.72
Projected Values Exceeding the Actual 2010 Mean						

- (b) Program 7.5.11 has 'Projected Mean' as the label on the last two columns in the DEFINE statements, the first should be 'Projected Median' (output is correct).
- (c) Program 7.6.2 Callout 1 should refer to Chapter Note 10 in 1.7 not Chapter Note 9 in 1.7.

8. Chapter 8

- (a) Output 8.3.1 and 8.3.2 reveal some inconsistencies about the application of the MONYY. format. If you look up the MONYY. format in the documentation, it shows it with capital letters for the month abbreviation. When used in procedures like PRINT, they are capital letters; however, here in PROC SGLOT they are in proper case.
- (b) Output 8.4.4 has too high of a lower limit for the y-axis, and the bottom of the high-low bars is cut off. The upper limit is also probably too high, creating extra white space. A span of 80 to 130 by 5 is a reasonable choice.