STA 311 Statistical Computing & Data Management

Instructor: Cheng Peng, Ph.D.
Department of Mathematics
West Chester University
West Chester, PA 19383

Office: 25 University Avenue, RM 111
Phone: 610.436.2369
Email: cpeng@wcupa.edu

Topics for This Week

- Subsetting and Splitting
- Subseeting Records
  - Partial Reading- FIRSTOBS, OBS
  - Conditional extraction: IF-THEN-ELSE, WHERE
- Subsetting Variables
  - Selecting designate variables: SELECT
  - Dropping designate variables: DROP
- Splitting
  - Conditional output of multiple data sets: IF-THEN-OUTPUT, SELECT-WHEN-OUTPUT-END

Extracting Records

Some Keywords: FIRSTOBS, OBS,

Extract cases beginning at row *i*:

```
DATA New-Dataset-Name (OPTIONS);
    SET Old-Dataset-Name (firstobs=i);
RUN;
```

Subsetting Datasets by Conditions

Some Keywords: IF-THEN-ELSE, WHERE, IN

Creating a subset that contains only records with a certain value

```
DATA New-Dataset-Name (OPTIONS);
    SET Old-Dataset-Name (OPTIONS);
    IF (insert conditions) THEN OUTPUT;
RUN;
```

Creating a subset that contains only records without a certain value

```
DATA New-Dataset-Name (OPTIONS);
    SET Old-Dataset-Name (OPTIONS);
    IF (insert conditions) THEN DELETE;
RUN;
```

Extracting Records

Extract the first *j* rows:

```
DATA New-Dataset-Name (OPTIONS);
    SET Old-Dataset-Name (obs=j);
RUN;
```

Extract rows *i* through *j*:

```
DATA New-Dataset-Name (OPTIONS);
    SET Old-Dataset-Name (firstobs=i obs=j);
RUN;
```

Extracting Variables

```
DATA auto2;
   SET auto;
   KEEP make mpg price;
RUN;
```

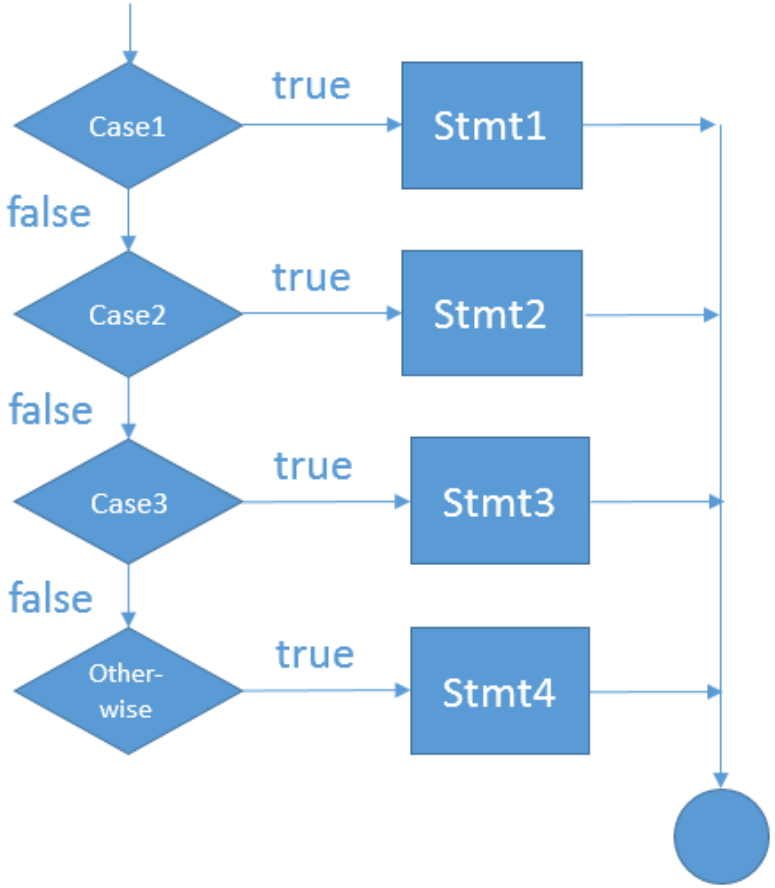
```
DATA auto3;
   SET auto;
   DROP rep78 displ gratio foreign;
RUN;
```

Splitting Data Sets

```
DATA New-Dataset-Name-1 (OPTIONS) New-Dataset-Name-2 (OPTIONS);
    SET Old-Dataset-Name (OPTIONS);
    IF (insert conditions for Dataset1) THEN OUTPUT New-Dataset-Name-1;
    IF (insert conditions for Dataset2) THEN OUTPUT New-Dataset-Name-2;
RUN;
```

```
DATA New-Dataset-Name-1 (OPTIONS) New-Dataset-Name-2 (OPTIONS);
    SET Old-Dataset-Name (OPTIONS);
    IF (insert conditions for Dataset1) THEN OUTPUT New-Dataset-Name-1;
    ELSE OUTPUT New-Dataset-Name-2;
RUN;
```

The SELECT statement in the SAS DATA step




```
/* example of using the SELECT statement */
DATA Non_smoker light_smkr mod_smkr
     Heavy_smkr VeryHeavy_smkr other_smkr;
SET sashelp.heart;
SELECT (Smoking_Status);
   WHEN ('Non-smoker')          OUTPUT Non_smoker;
   WHEN ('Light (1-5)')         OUTPUT light_smkr;
   WHEN ('Moderate (6-15)')     OUTPUT mod_smkr;
   WHEN ('Heavy (16-25)')       OUTPUT Heavy_smkr;
   WHEN ('Very Heavy (> 25)')   OUTPUT VeryHeavy_smkr;
   OTHERWISE                    OUTPUT other_smkr;
END;
RUN;
```

Caution: the values of the selected variable MUST be exhausted with WHEN-OTHERWISE clauses

WHERE and IF Statements

WHERE statement	Subsetting IF statement
Applied BEFORE observation read into PDV	Applied AFTER observation read into PDV
Can be used in many SAS procedures	Only used in the DATA step
Possible efficiency improvements	Reads and processes every record
Non-executable statement	Executable statement
Always applied at beginning of DATA step, regardless of where statement appears	Can be applied at any point in the DATA step, depending on where statement appears
Can use special operators such as CONTAINS, LIKE, and BETWEEN/AND	Can use automatic variables such as FIRST.BY, LAST.BY, and _N_
Can only access variables from input data set(s)	Can use newly created variables