**Chapter XXI – Reading Hierarchical Files**

Hierarchical consisting of a header record and one or more detail records

1. Creating One Observation per Detail Record

Necessary to distinguish between header and detail records, 所以在DATA step中RETAIN statement来retain header record。注意在每次读取时，都要keep相应的header record。

Eg:

**DATA** perm.people (**DROP**=type);

**INFILE** census;

**RETAIN** Address;

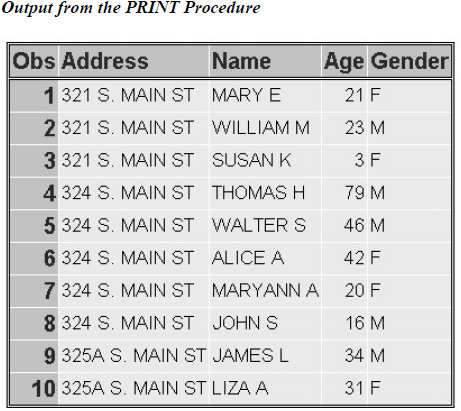
**INPUT** type $1. @;

**IT** type='H' **THEN INPUT** @3 address $15.;

**IF** type='P';

**INPUT** @3 Name $10. @13 Age 3. @16 Gender $1.;

**RUN**;



1. Creating One Observation per Header Record

主要作用是count一个header record有多少附属的detail record

* 重要code：
* **\_N\_** is an automatic variable whose value is the number of times the **DATA** step has begun to execute. The expression \_n\_ > 1 defines a condition where the **DATA** step has executed more than once. Use this expression in conjunction with the previous IF-THEN statement to check for these two conditions
* **END**= option

determine when the last record in the file is read so that you can then execute another explicit **OUTPUT** statement

* INFILE file-specification END=*variable*;

*variable* is a temporary numeric variable whose value is 0 until the last line is read and 1 after the last line is read.

* Eg:

**DATA** perm.residnts (**DROP**=type);

**INFILE** census **END**=last;

**RETAIN** Address;

**INPUT** type $1. @;

**IF** type='H' **THEN** do;

**IF** \_**N**\_ > 1 **THEN** **OUTPUT**;

Total=0;

**INPUT** address $ 3-17;

**END**;

**ELSE** **IF** type='P' **THEN** total+1;

**IF** last **THEN** output;

**RUN**；

1. Processing a DATA Step That Creates One Observation per Header Record



As the execution begins, \_N\_ is 1 and last is 0. Total is 0 because of the sum statement.





The condition N>1 is not true, so the OUTPUT statement is not executed. However, Total is assigned the value of 0 and the value for Address is read.





The value of last is still 0, so the OUTPUT statement is not executed. Control returns to the top of the

DATA step.



During the second iteration, the value of type is **'P'** and Total is incremented by 1. Again, the value of last is **0**, so control returns to the top of the DATA step.



During the fifth iteration, the value of type is **'H'** and \_N\_ is greater than 1, so the values for Address and Total are written to the data set as the first observation.



As the last record in the file is read, the variable last is set to **1**. Now that the condition for last is true, the values in the program data vector are written to the data set as the final observation.



练习

1. When you write a DATA step to create one observation per detail record you need to
2. distinguish between header and detail records.
3. keep the header record as a part of each observation until the next header record is encountered.
4. hold the current value of each record type so that the other values in the record can be read.
5. all of the above
6. Which SAS statement reads the value for code (in the first field), and then holds the value until an INPUT statement reads the remaining value in each observation in the same iteration of the DATA step?



1. input code $2. @;
2. input code $2. @@;
3. retain code;
4. none of the above
5. After the value for code is read in the sixth iteration, which illustration of the program data vector is correct?



**DATA** perm.produce (**DROP**=code);

**INFILE** orders;

**RETAIN** Vegetable;

**INPUT** code $1. @;

**IF** code='H' **THEN** **INPUT** @3 vegetable $6.;

**IF** code='P';

**INPUT** @3 Variety : $10. @15 Supplier : $15.;

**RUN**;

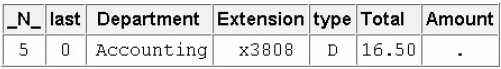
**PROC** **PRINT** data=perm.produce;

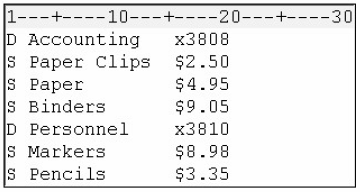
**RUN**;

1. 
2. 
3. 
4. 
5. Which SAS statement indicates that several other statements should be executed when Record has a value of A?



1. **IF** record='A' **THEN** **DO**;
2. **IF** record=A **THEN** **DO**;
3. **IF** record='A' **THEN**;
4. **IF** record=A **THEN**;
5. Based on the values in the program data vector, what happens next?





**DATA** work.supplies (**DROP**=type amount);

**INFILE** orders **END**=**LAST**;

**RETAIN** Department Extension;

**INPUT** type $1. @;

**IF** type='D' **THEN** **DO**;

**IF** \_**N**\_ > 1 **THEN** **OUTPUT**;

Total=0;

**INPUT** @3 department $10. @16 extension $5.;

**END**;

**ELSE** **IF** type='S' **THEN** **DO**;

**INPUT** @16 Amount comma5.;

total+amount;

**IF** last **THEN** **OUTPUT**;

**END**;

**RUN**;

1. All the values in the program data vector are written to the data set as the first observation.
2. The values for Department, Total, and Extension are written to the data set as the first observation.
3. The values for Department, Total, and Extension are written to the data set as the fourth observation.
4. The value of last changes to 1.