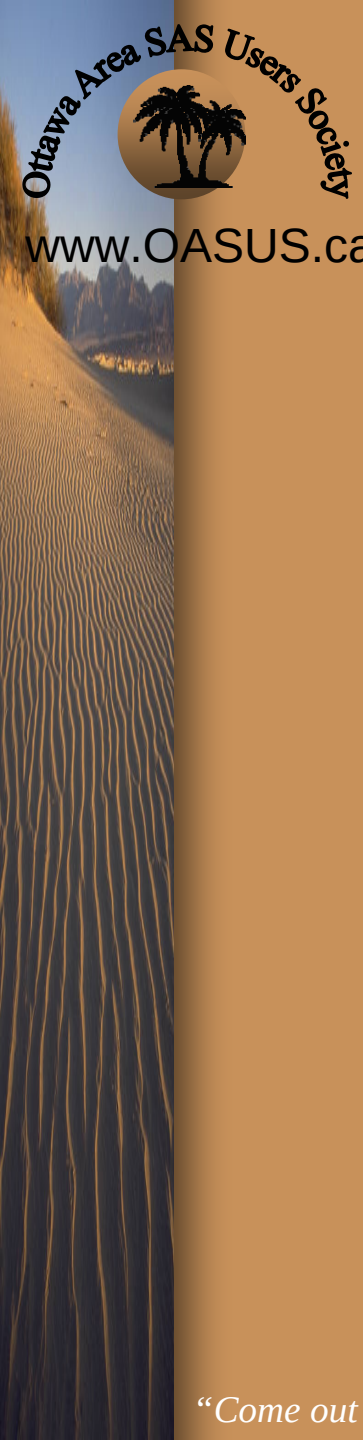


www.OASUS.ca

Look Before You Leap

Leonard Landry
SCAN, Statistics Canada

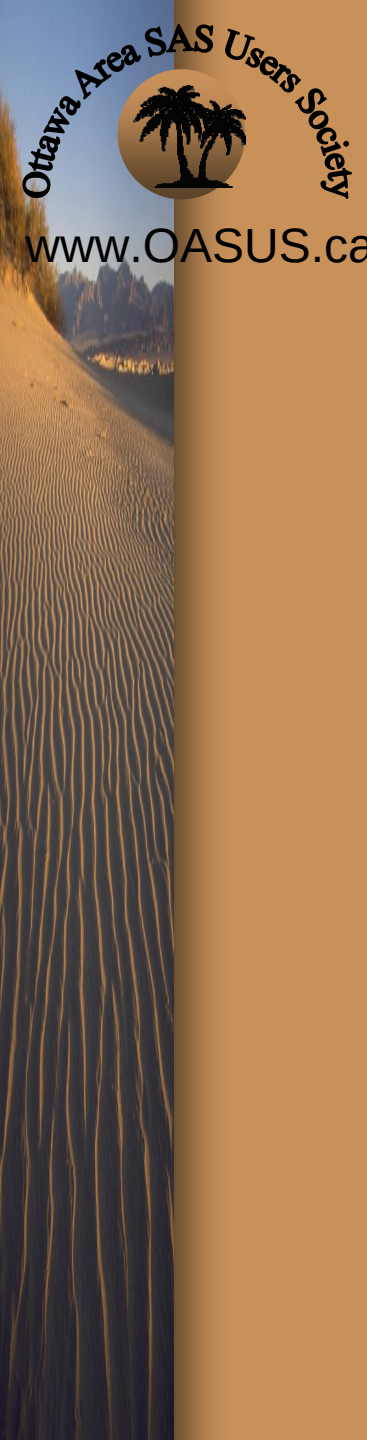
"Come out of the desert of ignorance to the OASUS of knowledge"

Where are we going

- A technique to achieve a Look-Ahead Capability
  - How and why it works
  - Two variations of implementing it
  - A practical example
  - Conclusions

Theory

- When SAS reads a SAS dataset sequentially it maintains a pointer to the current observation.
- Each reference to a SAS Dataset will generate it's own pointer.
- We can read the same SAS dataset twice in a SAS data step with the pointer in the second read maintained one observation ahead of the first.



Sample SAS Dataset

OBS	VAR1	VAR2
1	1	A
2	2	B
3	3	C
4	4	D
5	5	E

Example 1

```
data test2;
   set test;
   set test(Firstobs=2 keep=var2
                rename=(var2=nextvar2));
run;
```

Example 1

NOTE: There were 5 observations read from the data set WORK.TEST.
NOTE: There were 4 observations read from the data set WORK.TEST.
NOTE: The data set WORK.TEST2 has 4 observations and 3 variables.
NOTE: DATA statement used (Total process time):
real time 0.01 seconds
cpu time 0.03 seconds

OBS	VAR1	VAR2	NEXTVAR2
1	1	A	B
2	2	B	C
3	3	C	D
4	4	D	E

Example 1

```
data test2;
   set test;
   If eof=0
      then set test(firstobs=2 keep=var2
                rename=(var2=nextvar2)) end=eof;
run;
```

Example 1

NOTE: There were 5 observations read from the data set WORK.TEST.
NOTE: There were 4 observations read from the data set WORK.TEST.
NOTE: The data set WORK.TEST2 has 5 observations and 3 variables.
NOTE: DATA statement used (Total process time):
real time 0.00 seconds
cpu time 0.00 seconds

OBS	VAR1	VAR2	NEXTVAR2
1	1	A	B
2	2	B	C
3	3	C	D
4	4	D	E
5	5	E	E

Example 1

```
data test2;
   set test;
   if eof=0
      then set test(firstobs=2 keep=var2
                    rename=(var2=nextvar2)) end=eof;
      else nextvar2=' ';
run;
```

Example 1

NOTE: There were 5 observations read from the data set WORK.TEST.
NOTE: There were 4 observations read from the data set WORK.TEST.
NOTE: The data set WORK.TEST2 has 5 observations and 3 variables.
NOTE: DATA statement used (Total process time):
real time 0.00 seconds
cpu time 0.00 seconds

OBS	VAR1	VAR2	NEXTVAR2
1	1	A	B
2	2	B	C
3	3	C	D
4	4	D	E
5	5	E	

Example 2

```
Data test2;
  merge test test(firstobs=2 keep=var2
                  rename=(var2=nextvar2));
run;
```

Example 2

NOTE: There were 5 observations read from the data set WORK.TEST.
NOTE: There were 4 observations read from the data set WORK.TEST.
NOTE: The data set WORK.TEST2 has 5 observations and 3 variables.
NOTE: DATA statement used (Total process time):
real time 0.01 seconds
cpu time 0.03 seconds

OBS	VAR1	VAR2	NEXTVAR2
1	1	A	B
2	2	B	C
3	3	C	D
4	4	D	E
5	5	E	

Example 3

- Longitudinal File (1983 – 2007)
- Data on individual workers including job earnings
- Worker ID is consistent across all years.
- File contains data for a given worker in each year the worker had earned income
- We want to know how many worker had income in one year but not in the next year

Example 3

```
proc sort  nodupkey data=lwf.lwf8307(keep=rectype worker_id year age
     sex)   out=wagegap;
   where rectype < 3 & sex>0 & age>0;
   by worker_id year;

/* find all workers with a wage gap from LWF file */

DATA wagegap2(keep=year count_of_wagegaps);
  array wagegap_counts(1983:2006) 8. _temporary_;
  merge wagegap wagegap(firstobs=2 keep=worker_id year
        rename=(worker_id=nextworker year=nextyear)) end=end1;
  if year < 2007 & ((nextworker = worker_id & nextyear ne year + 1)
                  | nextworker ne worker_id)
     then wagegap_counts(year)+1;
  if end1  then do year=1983 to 2006;
                        count_of_wagegaps = wagegap_counts(year);
                        output;
                    end;
```

Example3

Year	count_of_wagegaps
1983	94735
1984	90315
1985	92683
1986	90393
1987	92207
1988	93155
1989	106048
1990	124425
1991	128399
1992	122040
1993	112479
1994	107832
1995	112360
1996	105056
1997	104910
1998	97789
1999	93840
2000	104369
2001	108626
2002	102389
2003	113653
2004	104981
2005	105808
2006	103755

Conclusions

- Both variations of the “look ahead” technique demonstrated here work
- The second technique, merging a dataset to itself, is simpler and easier to use
- The “look ahead” facility can also be used to look ahead by more than 1 record by varying the value in the FIRSTOBS option

Ottawa Area SAS Users Society

www.OASUS.ca

Questions / Comments

Statistics Canada | Statistique Canada

Leonard Landry
Database Manager
Social Analysis Division
R.H. Coats Building, 24th Floor, Section E

Ottawa, Ontario, Canada K1A 0T6
(613) 951-1284 Fax (613) 951-5403

Leonard.landry@statcan.gc.ca

Canada