

## Longitudinal Data Techniques:

### Looking Across Observations

Ronald Cody, Ed.D., Robert Wood Johnson Medical School

## A Typical Longitudinal Data Set

PATNO	DATE	DOB	HR	SBP	DBP
001	10/21/1997	10/21/1946	48	128	74
001	02/01/1998		44	126	70
001	11/04/1998		52	130	76
001	11/07/1998		54	132	78
003	11/11/1998	09/08/1955	52	140	80
007	04/04/1998				
008	03/22/1998	02/08/1980	58	144	72
008	04/21/1998		66	144	74
012	05/06/1998		80	120	80
013	11/11/1998	11/09/1930	100	180	108
013	11/18/1998		90	170	98
123	01/28/1998	01/01/1944	80	180	96
123	05/04/1998		80	178	90

## Common Data Processing Requirements

- Extract the first or last visit for a subject
- Copy information from the first visit to subsequent visits
- Compute intra-patient statistics such as mean, median, minimum, or maximum

## Common Data Processing Requirements

- Count the number of visits per patient
- Compute differences of variables between visits
- Make decisions on the current visit, based on information from a future visit

## Selecting the First (or last) Visit for Each Patient

```
PROC SORT DATA=LABS;  
  BY PATNO DATE;  
RUN;  
DATA FIRST;  
  SET LABS;  
  BY PATNO;  
  IF FIRST.PATNO;  
RUN;  
DATA LAST;  
  SET LABS;  
  BY PATNO;  
  IF LAST.PATNO;  
RUN;
```

### Listing of FIRST

PATNO	DATE	DOB	HR	SBP	DBP
001	10/21/1997	10/21/1946	48	128	74
003	11/11/1998	09/08/1955	52	140	80
007	04/04/1998				
008	03/22/1998	02/08/1980	58	144	72
012	05/06/1998		80	120	80
013	11/11/1998	11/09/1930	100	180	108
123	01/28/1998	01/01/1944	80	180	96
354	04/12/1998	07/07/1955	90	210	110
554	06/08/1998	09/12/1944	48	108	66
888	01/01/1998	03/14/1922	46	110	68

## Adding DOB to the Second Through the Last Observation for Each Patient

```
PROC SORT DATA=LABS;  
  BY PATNO DATE;  
RUN;  
DATA LAB2;  
  SET LAB;  
  BY PATNO;  
  RETAIN OLD_DOB;  
  IF FIRST.PATNO THEN OLD_DOB = DOB;  
  ELSE DOB = OLD_DOB;  
RUN;
```

## DOB Added to Observations

PATNO	DATE	DOB	HR	SBP	DBP
001	10/21/1997	10/21/1946	48	128	74
001	02/01/1998	10/21/1946	44	126	70
001	11/04/1998	10/21/1946	52	130	76
001	11/07/1998	10/21/1946	54	132	78
003	11/11/1998	09/08/1955	52	140	80
007	04/04/1998				
008	03/22/1998	02/08/1980	58	144	72
008	04/21/1998	02/08/1980	66	144	74
012	05/06/1998		80	120	80
013	11/11/1998	11/09/1930	100	180	108
013	11/18/1998	11/09/1930	90	170	98
123	01/28/1998	01/01/1944	80	180	96
123	05/04/1998	01/01/1944	80	178	90

## Computing Differences Between Observations (using RETAIN)

```
*DIFFERENCE BETWEEN FIRST AND LAST VISIT;
DATA DIFFERENCE; /*NOT USING V8? */
SET LAB2;
BY PATNO;
*REMOVE PATIENTS WITH ONE VISIT;
IF FIRST.PATNO AND LAST.PATNO THEN DELETE;
RETAIN R_HR R_SBP R_DBP;
IF FIRST.PATNO THEN DO;
    R_HR = HR;
    R_SBP = SBP;
    R_DBP = DBP;
END;
```

(continued)

## Computing Differences Between Observations (using RETAIN)

(continued)

```
IF LAST.PATNO THEN DO;
    DIFF_HR = HR - R_HR;
    DIFF_SBP = SBP - R_SBP;
    DIFF_DBP = DBP - R_DBP;
    OUTPUT;
END;
DROP R_ ;
RUN;
```

## Computing Differences Between Observations (using RETAIN)

Listing of data set difference

P	A	D	S	D	D	D	D	D
O	b	T	O	H	B	B	H	S
s	O	E	B	R	P	P	R	P
1	001	11/07/1998	10/21/1946	54	132	78	6	4
2	008	04/21/1998	02/08/1980	66	144	74	8	0
3	013	11/18/1998	11/09/1930	90	170	98	-10	-10
4	123	05/04/1998	01/01/1944	80	178	90	0	-2

## Computing Differences Between the First and Last Visit (using LAG)

```
DATA DIFF3;
SET LABS2;
BY PATNO;
*Remove patients with one visit;
IF FIRST.PATNO AND LAST.PATNO THEN DELETE;
IF FIRST.PATNO OR LAST.PATNO THEN DO;
    DIFF_HR = HR - LAG(HR);
    DIFF_SBP = SBP - LAG(SBP);
    DIFF_DBP = SBP - LAG(DBP);
END;
IF LAST.PATNO THEN OUTPUT;
RUN;
```

Note: About the only time I ever executed a LAG function conditionally (on purpose)

## Computing Differences Between Observations (using the LAG function)

```
*DIFFERENCES BETWEEN ALL VISITS;
DATA DIFF2;
SET LAB2;
BY PATNO;
DIFF_HR = HR - LAG(HR);
*Alternative (below) using the DIF function;
DIFF_SBP = DIF(SBP);
DIFF_DBP = DIF(DBP);
IF NOT FIRST.PATNO THEN OUTPUT;
RUN;
```

## Computing Differences Between Observations (using the LAG function)

Listing of data set diff2

```

P          D          D          S          D          D          D          D          D
A          A          O          H          B          B          H          S          I
b          T          O          B          R          P          R          B          F
s          E          B          R          P          P          R          P          F

```

	P	A	D	O	H	B	R	P	D	S	I	D
1	001	02/01/1998	10/21/1946	44	126	70	-4	-2	-4			
2	001	11/04/1998	10/21/1946	52	130	76	8	4	6			
3	001	11/07/1998	10/21/1946	54	132	78	2	2	2			
4	008	04/21/1998	02/08/1980	66	144	74	8	0	2			
5	013	11/18/1998	11/09/1930	90	170	98	-10	-10	-10			
6	123	05/04/1998	01/01/1944	80	178	90	0	-2	-6			

## Counting the Number of Observations in Each BY Group

```

DATA COUNT_IT;
SET LABS2(KEEP=PATNO);
BY PATNO;
IF FIRST.PATNO THEN N_VISITS = 1;
ELSE N_VISITS + 1;
IF LAST.PATNO THEN OUTPUT;
RUN;

```

Listing of data set COUNT\_IT

Obs	PATNO	N_VISITS
1	001	4
2	003	1
3	007	1
4	008	2
5	012	1
6	013	2
7	123	2

## Using PROC FREQ to Output a Data Set Containing Counts

```

PROC FREQ DATA=LABS2 NOPRINT;
TABLES PATNO /
OUT=COUNTS(KEEP=PATNO COUNT
RENAME=(COUNT=N_VISITS));
RUN;

```

Listing of data set COUNTS

Obs	PATNO	N_VISITS
1	001	4
2	003	1
3	007	1
4	008	2
5	012	1
6	013	2
7	123	2

## Creating Summary Data Sets Using PROC MEANS

```

PROC MEANS DATA=LABS2 NWAY NOPRINT;
CLASS PATNO;
VAR HR SBP DBP;
OUTPUT
OUT = SUMS(DROP=_TYPE_
RENAME=( _FREQ_ = N_VISITS))
N = N_HR N_SBP N_DBP
MEAN = M_HR M_SBP M_DBP;
RUN;

```

## Creating Summary Data Sets Using PROC MEANS (OUTPUT)

Listing of data set SUMS

Obs	PATNO	N_VISITS	N_HR	N_SBP	N_DBP	M_HR	M_SBP	M_DBP
1	001	4	4	4	4	49.5	129	74.5
2	003	1	1	1	1	52.0	140	80.0
3	007	1	0	0	0	.	.	.
4	008	2	2	2	2	62.0	144	73.0
5	012	1	1	1	1	80.0	120	80.0
6	013	2	2	2	2	95.0	175	103.0
7	123	2	2	2	2	80.0	179	93.0

## Selecting All Patients with "n" Visits (using PROC FREQ)

```

PROC FREQ DATA=LABS2 NOPRINT;
TABLES PATNO /
OUT = COUNTS(KEEP=PATNO COUNT
RENAME=(COUNT=N_VISITS)
WHERE=(N_VISITS = 2));
RUN;
DATA TWO_VISITS;
MERGE COUNTS(IN=IN_COUNT) LABS2;
BY PATNO;
IF IN_COUNT;
RUN;

```

## Selecting All Patients with "n" Visits (OUTPUT)

Listing of data set TWO\_VISITS

Obs	PATNO	N_VISITS	DATE	DOB	HR	SBP	DBP
1	008	2	03/22/1998	02/08/1980	58	144	72
2	008	2	04/21/1998	02/08/1980	66	144	74
3	013	2	11/11/1998	11/09/1930	100	180	108
4	013	2	11/18/1998	11/09/1930	90	170	98
5	123	2	01/28/1998	01/01/1944	80	180	96
6	123	2	05/04/1998	01/01/1944	80	178	90

## Using PROC SQL and a Macro Variable to Select Observations

```
PROC SQL NOPRINT;
  SELECT QUOTE(PATNO)
  INTO :DUP_LIST SEPARATED BY " "
  FROM LABS2
  GROUP BY PATNO
  HAVING FREQ(PATNO) EQ 2;
QUIT;
PROC PRINT DATA=LABS2;
  WHERE PATNO IN (&DUP_LIST);
  TITLE "Using SQL and a Macro Variable";
RUN;
```

```
DUP_LIST '008' '013' '123'
```

## Selecting All Patients with "n" Visits (using a Data Step)

```
DATA TWO;
  SET LABS(KEEP=PATNO);
  BY PATNO;
  IF FIRST.PATNO THEN N = 1;
  ELSE N + 1;
  IF LAST.PATNO AND N = 2 THEN OUTPUT;
RUN;
DATA TWO_VISITS;
  MERGE TWO(IN=IN_COUNT) LABS2;
  BY PATNO;
  IF IN_COUNT;
RUN;
```

## Looking Ahead Using Multiple SET Statements

```
DATA DOC;
  INPUT @1 PATNO $3.
  @5 VISIT MMDDYY10.
  @16 DOCTOR $3.;
  FORMAT VISIT MMDDYY10.;
  DATALINES;
001 10/21/1998 ABC
001 10/29/1998 XYZ ← Failure by ABC
001 12/12/1998 QED
002 01/01/1998 ABC
003 02/13/1998 QED
003 04/15/1998 MAD
005 05/06/1998 XYZ
005 05/08/1998 QED ← Failure by XYZ
;
```

## Looking Ahead Using Multiple SET Statements

```
PROC SORT DATA=DOC;
  BY PATNO VISIT;
RUN;
DATA FAILURES;
  SET DOC;
  BY PATNO;
  SET DOC (FIRSTOBS = 2
  KEEP = VISIT
  RENAME = (VISIT = NEXT_VISIT));
  IF NOT LAST.PATNO AND
  (NEXT_VISIT - VISIT) LT 30 THEN OUTPUT;
  KEEP PATNO VISIT NEXT_VISIT DOCTOR;
RUN;
```

```
001 10/21/1998 ABC
001 10/29/1998 XYZ
001 12/12/1998 QED
002 01/01/1998 ABC
003 02/13/1998 QED
003 04/15/1998 MAD
005 05/06/1998 XYZ
005 05/08/1998 QED
```

## Looking Ahead Using Multiple SET Statements

```
PROC SORT DATA=DOC;
  BY PATNO VISIT;
RUN;
DATA FAILURES;
  SET DOC;
  BY PATNO;
  SET DOC (FIRSTOBS = 2
  KEEP = VISIT
  RENAME = (VISIT = NEXT_VISIT));
  IF NOT LAST.PATNO AND
  (NEXT_VISIT - VISIT) LT 30 THEN OUTPUT;
  KEEP PATNO VISIT NEXT_VISIT DOCTOR;
RUN;
```

Listing of data set FAILURES

Obs	PATNO	VISIT	DOCTOR	NEXT_VISIT
1	001	10/21/1998	ABC	10/29/1998
2	005	05/06/1998	XYZ	05/08/1998