Handbook of SAS®
Data Step Programming

Execution Phase of Program 4.1

Arthur Li

Execution Phase of Program 4.1

```
data sas4_2(drop=score);
     set sas4_1;
     by id;
     if first.id then total = 0;
     total + score;
     if last.id;
run;
```

SAS4_1:

	ID	SCORE
1	A01	3
2	A01	3
3	A01	2
4	A02	4
5	A02	2

PDV

_N_	D	FIRST.ID	D	LAST.ID	D	ID	K	SCORE	D	TOTAL	K
1		1		1				.		0	

1st Iteration:

- _N_ ←1

ERROR is not shown for purpose of simplicity

Execution Phase of Program 4.1

```
data sas4_2(drop=score);
     set sas4_1;
     by id;
     if first.id then total = 0;
     total + score;
     if last.id;
run;
```

SAS4_1:

	ID	SCORE
1	A01	3
2	A01	3
3	A01	2
4	A02	4
5	A02	2

PDV

_N_	D	FIRST.ID	D	LAST.ID	D	ID	K	SCORE	D	TOTAL	K
1		1		1				.		0	

1st Iteration:

- _N_ ←1
- FIRST.ID ← 1, LAST.ID ← 1

Execution Phase of Program 4.1

```
data sas4_2(drop=score);
     set sas4_1;
     by id;
     if first.id then total = 0;
     total + score;
     if last.id;
run;
```

SAS4_1:

	ID	SCORE
1	A01	3
2	A01	3
3	A01	2
4	A02	4
5	A02	2

PDV

_N_	D	FIRST.ID	D	LAST.ID	D	ID	K	SCORE	D	TOTAL	K
1		1		1				.		0	

1st Iteration:

- _N_ ←1
- FIRST.ID ← 1, LAST.ID ← 1
- ID, Score ← missing

Execution Phase of Program 4.1

```
data sas4_2(drop=score);
     set sas4_1;
     by id;
     if first.id then total = 0;
     total + score;
     if last.id;
run;
```

SAS4_1:

	ID	SCORE
1	A01	3
2	A01	3
3	A01	2
4	A02	4
5	A02	2

PDV

_N_	D	FIRST.ID	D	LAST.ID	D	ID	K	SCORE	D	TOTAL	K
1		1		1				.		0	

1st Iteration:

- _N_ ←1
- FIRST.ID ← 1, LAST.ID ← 1
- ID, Score ← missing
- TOTAL ← 0 because of the SUM statement

Execution Phase of Program 4.1

```
data sas4_2(drop=score);
   set sas4_1;
   by id;
   if first.id then total = 0;
   total + score;
   if last.id;
run;
```

SAS4_1:

	ID	SCORE
1	A01	3
2	A01	3
3	A01	2
4	A02	4
5	A02	2

PDV

_N_ D	FIRST.ID D	LAST.ID D	ID K	SCORE D	TOTAL K
1	1	0	A01	3	0

1st Iteration:

- 1st observation → PDV

Execution Phase of Program 4.1

```
data sas4_2(drop=score);
   set sas4_1;
   by id;
   if first.id then total = 0;
   total + score;
   if last.id;
run;
```

SAS4_1:

	ID	SCORE
1	A01	3
2	A01	3
3	A01	2
4	A02	4
5	A02	2

PDV

_N_	D	FIRST.ID	D	LAST.ID	D	ID	K	SCORE	D	TOTAL	K
1		1		0		A01		3		0	

1st Iteration:

- 1st observation → PDV
- FIRST.ID ← 1 and LAST.ID ← 0

Execution Phase of Program 4.1

```
data sas4_2(drop=score);
     set sas4_1;
     by id;
     if first.id then total = 0;
     total + score;
     if last.id;
run;
```

SAS4_1:

	ID	SCORE
1	A01	3
2	A01	3
3	A01	2
4	A02	4
5	A02	2

PDV

_N_	D	FIRST.ID	D	LAST.ID	D	ID	K	SCORE	D	TOTAL	K
1		1		0		A01		3		0	

1st Iteration:

- BY statement is a declarative statement

Execution Phase of Program 4.1

```
data sas4_2(drop=score);
     set sas4_1;
     by id;
     if first.id then total = 0;
     total + score;
     if last.id;
run;
```

SAS4_1:

	ID	SCORE
1	A01	3
2	A01	3
3	A01	2
4	A02	4
5	A02	2

PDV

_N_	D	FIRST.ID	D	LAST.ID	D	ID	K	SCORE	D	TOTAL	K
1		1		0		A01		3		0	

1st Iteration:

- FIRST.ID = 1: TOTAL ← 0

Execution Phase of Program 4.1

```
data sas4_2(drop=score);
     set sas4_1;
     by id;
     if first.id then total = 0;
     total + score;
     if last.id;
run;
```

SAS4_1:

	ID	SCORE
1	A01	3
2	A01	3
3	A01	2
4	A02	4
5	A02	2

PDV

_N_	D	FIRST.ID	D	LAST.ID	D	ID	K	SCORE	D	TOTAL	K
1		1		0		A01		3		3	

1st Iteration:

- TOTAL is calculated

Execution Phase of Program 4.1

```
data sas4_2(drop=score);
     set sas4_1;
     by id;
     if first.id then total = 0;
     total + score;
  ➔  if last.id;
run;
```

SAS4_1:

	ID	SCORE
1	A01	3
2	A01	3
3	A01	2
4	A02	4
5	A02	2

PDV

_N_	D	FIRST.ID	D	LAST.ID	D	ID	K	SCORE	D	TOTAL	K
1		1		0		A01		3		3	

1st Iteration:

- Since LAST.ID ≠ 1, (the subsetting IF statement is false), no further statements are processed for the current observation. SAS immediately returns to the beginning of the DATA step

Execution Phase of Program 4.1

```
data sas4_2(drop=score);
     set sas4_1;
     by id;
     if first.id then total = 0;
     total + score;
     if last.id;
run;
```

SAS4_1:

	ID	SCORE
1	A01	3
2	A01	3
3	A01	2
4	A02	4
5	A02	2

PDV

_N_	D	FIRST.ID	D	LAST.ID	D	ID	K	SCORE	D	TOTAL	K
2		1		0		A01		3		3	

2nd Iteration:

- _N_ ↑ 2

Execution Phase of Program 4.1

```
data sas4_2(drop=score);
     set sas4_1;
     by id;
     if first.id then total = 0;
     total + score;
     if last.id;
run;
```

SAS4_1:

	ID	SCORE
1	A01	3
2	A01	3
3	A01	2
4	A02	4
5	A02	2

PDV

_N_	D	FIRST.ID	D	LAST.ID	D	ID	K	SCORE	D	TOTAL	K
2		1		0		A01		3		3	

2nd Iteration:

- _N_ ↑ 2
- FIRST.ID & LAST.ID are retained (automatic variables)

Execution Phase of Program 4.1

```
data sas4_2(drop=score);
     set sas4_1;
     by id;
     if first.id then total = 0;
     total + score;
     if last.id;
run;
```

SAS4_1:

	ID	SCORE
1	A01	3
2	A01	3
3	A01	2
4	A02	4
5	A02	2

PDV

_N_	D	FIRST.ID	D	LAST.ID	D	ID	K	SCORE	D	TOTAL	K
2		1		0		A01		3		3	

2nd Iteration:

- _N_ ↑ 2
- FIRST.ID & LAST.ID are retained (automatic variables)
- ID & SCORE are retained (read from input data)

Execution Phase of Program 4.1

```
data sas4_2(drop=score);
     set sas4_1;
     by id;
     if first.id then total = 0;
     total + score;
     if last.id;
run;
```

SAS4_1:

	ID	SCORE
1	A01	3
2	A01	3
3	A01	2
4	A02	4
5	A02	2

PDV

_N_	D	FIRST.ID	D	LAST.ID	D	ID	K	SCORE	D	TOTAL	K
2		1		0		A01		3		3	

2nd Iteration:

- _N_ ↑ 2
- FIRST.ID & LAST.ID are retained (automatic variables)
- ID & SCORE are retained (read from input data)
- TOTAL is retained (SUM statement)

Execution Phase of Program 4.1

```
data sas4_2(drop=score);
   set sas4_1;
     by id;
     if first.id then total = 0;
     total + score;
     if last.id;
run;
```

SAS4_1:

	ID	SCORE
1	A01	3
2	A01	3
3	A01	2
4	A02	4
5	A02	2

PDV

_N_ D	FIRST.ID D	LAST.ID D	ID K	SCORE D	TOTAL K
2	0	0	A01	3	3

2nd Iteration:

- 2nd observation → PDV

Execution Phase of Program 4.1

```
data sas4_2(drop=score);
   set sas4_1;
     by id;
     if first.id then total = 0;
     total + score;
     if last.id;
run;
```

SAS4_1:

	ID	SCORE
1	A01	3
2	A01	3
3	A01	2
4	A02	4
5	A02	2

PDV

_N_	D	FIRST.ID	D	LAST.ID	D	ID	K	SCORE	D	TOTAL	K
2		0		0		A01		3		3	

2nd Iteration:

- 2nd observation → PDV
- Not the first observation for A01: FIRST.ID ←0
- Not the last observation for A01: LAST.ID ← 0

Execution Phase of Program 4.1

```
data sas4_2(drop=score);
     set sas4_1;
     by id;
     if first.id then total = 0;
     total + score;
     if last.id;
run;
```

SAS4_1:

	ID	SCORE
1	A01	3
2	A01	3
3	A01	2
4	A02	4
5	A02	2

PDV

_N_	D	FIRST.ID	D	LAST.ID	D	ID	K	SCORE	D	TOTAL	K
2		0		0		A01		3		3	

2nd Iteration:

- FIRST.ID ≠ 1: no execution

Execution Phase of Program 4.1

```
data sas4_2(drop=score);
     set sas4_1;
     by id;
     if first.id then total = 0;
  ➔  total + score;
     if last.id;
run;
```

SAS4_1:

	ID	SCORE
1	A01	3
2	A01	3
3	A01	2
4	A02	4
5	A02	2

PDV

_N_ D	FIRST.ID D	LAST.ID D	ID K	SCORE D	TOTAL K
2	0	0	A01	3	6

2nd Iteration:

- TOTAL is calculated

Execution Phase of Program 4.1

```
data sas4_2(drop=score);
     set sas4_1;
     by id;
     if first.id then total = 0;
     total + score;
  ➔  if last.id;
run;
```

SAS4_1:

	ID	SCORE
1	A01	3
2	A01	3
3	A01	2
4	A02	4
5	A02	2

PDV

_N_ D	FIRST.ID D	LAST.ID D	ID K	SCORE D	TOTAL K
2	0	0	A01	3	6

2nd Iteration:

- Since LAST.ID ≠ 1 (the subsetting IF statement is false), SAS immediately returns to the beginning of the DATA step

Execution Phase of Program 4.1

```
data sas4_2(drop=score);
     set sas4_1;
     by id;
     if first.id then total = 0;
     total + score;
     if last.id;
run;
```

SAS4_1:

	ID	SCORE
1	A01	3
2	A01	3
3	A01	2
4	A02	4
5	A02	2

PDV

_N_	D	FIRST.ID	D	LAST.ID	D	ID	K	SCORE	D	TOTAL	K
3		0		0		A01		3		6	

3rd Iteration:

- _N_ ↑3
- The values for the rest of the variables are retained

Execution Phase of Program 4.1

```
data sas4_2(drop=score);
   set sas4_1;
     by id;
     if first.id then total = 0;
     total + score;
     if last.id;
run;
```

SAS4_1:

	ID	SCORE
1	A01	3
2	A01	3
3	A01	2
4	A02	4
5	A02	2

PDV

_N_	D	FIRST.ID	D	LAST.ID	D	ID	K	SCORE	D	TOTAL	K
3		0		1		A01		2		6	

3rd Iteration:

- 3rd observation →PDV

Execution Phase of Program 4.1

```
data sas4_2(drop=score);
   set sas4_1;
     by id;
     if first.id then total = 0;
     total + score;
     if last.id;
run;
```

SAS4_1:

	ID	SCORE
1	A01	3
2	A01	3
3	A01	2
4	A02	4
5	A02	2

PDV

_N_	D	FIRST.ID	D	LAST.ID	D	ID	K	SCORE	D	TOTAL	K
3		0		1		A01		2		6	

3rd Iteration:

- 3rd observation →PDV
- Not the first observation: FIRST.ID ← 0
- Last observation for A01: LAST.ID ← 1

Execution Phase of Program 4.1

```
data sas4_2(drop=score);
     set sas4_1;
     by id;
     if first.id then total = 0;
     total + score;
     if last.id;
run;
```

SAS4_1:

	ID	SCORE
1	A01	3
2	A01	3
3	A01	2
4	A02	4
5	A02	2

PDV

_N_	D	FIRST.ID	D	LAST.ID	D	ID	K	SCORE	D	TOTAL	K
3		0		1		A01		2		6	

3rd Iteration:

- FIRST.ID ≠ 1: no execution

Execution Phase of Program 4.1

```
data sas4_2(drop=score);
     set sas4_1;
     by id;
     if first.id then total = 0;
     total + score;
     if last.id;
run;
```

SAS4_1:

	ID	SCORE
1	A01	3
2	A01	3
3	A01	2
4	A02	4
5	A02	2

PDV

_N_ D	FIRST.ID D	LAST.ID D	ID K	SCORE D	TOTAL K
3	0	1	A01	2	8

3rd Iteration:

- TOTAL is calculated

Execution Phase of Program 4.1

```
data sas4_2(drop=score);
     set sas4_1;
     by id;
     if first.id then total = 0;
     total + score;
  ➔  if last.id;
run;
```

SAS4_1:

	ID	SCORE
1	A01	3
2	A01	3
3	A01	2
4	A02	4
5	A02	2

PDV

_N_	D	FIRST.ID	D	LAST.ID	D	ID	K	SCORE	D	TOTAL	K
3		0		1		A01		2		8	

3rd Iteration:

- Since LAST.ID = 1 (the subsetting IF statement is true), SAS continues to execute the remaining statements in the DATA step

Execution Phase of Program 4.1

```
data sas4_2(drop=score);
     set sas4_1;
     by id;
     if first.id then total = 0;
     total + score;
     if last.id;
run;
```

SAS4_1:

	ID	SCORE
1	A01	3
2	A01	3
3	A01	2
4	A02	4
5	A02	2

PDV

_N_	D	FIRST.ID	D	LAST.ID	D	ID	K	SCORE	D	TOTAL	K
3		0		1		A01		2		8	

3rd Iteration:

- SAS reaches the end of the 3rd iteration
  - The implicit OUTPUT statement copies ID and TOTAL in the PDV to the output data set
  - SAS returns to the beginning of the DATA step to begin the 4th iteration

SAS4_2:

	ID	TOTAL
1	A01	8

Execution Phase of Program 4.1

```
data sas4_2(drop=score);
     set sas4_1;
     by id;
     if first.id then total = 0;
     total + score;
     if last.id;
run;
```

SAS4_1:

	ID	SCORE
1	A01	3
2	A01	3
3	A01	2
4	A02	4
5	A02	2

PDV

_N_	D	FIRST.ID	D	LAST.ID	D	ID	K	SCORE	D	TOTAL	K
4		0		1		A01		2		8	

4th Iteration:

- _N_ ↑ 4
- The values for the remaining variables are retained

SAS4_2:

	ID	TOTAL
1	A01	8

Execution Phase of Program 4.1

```
data sas4_2(drop=score);
   set sas4_1;
     by id;
     if first.id then total = 0;
     total + score;
     if last.id;
run;
```

SAS4_1:

	ID	SCORE
1	A01	3
2	A01	3
3	A01	2
4	A02	4
5	A02	2

PDV

_N_	D	FIRST.ID	D	LAST.ID	D	ID	K	SCORE	D	TOTAL	K
4		1		0		A02		4		8	

4th Iteration:

- 4th observation → PDV

SAS4_2:

	ID	TOTAL
1	A01	8

Execution Phase of Program 4.1

```
data sas4_2(drop=score);
   set sas4_1;
     by id;
     if first.id then total = 0;
     total + score;
     if last.id;
run;
```

SAS4_1:

	ID	SCORE
1	A01	3
2	A01	3
3	A01	2
4	A02	4
5	A02	2

PDV

_N_	D	FIRST.ID	D	LAST.ID	D	ID	K	SCORE	D	TOTAL	K
4		1		0		A02		4		8	

4th Iteration:

- 4th observation → PDV
- FIRST.ID ←1
- LAST.ID ← 0

SAS4_2:

	ID	TOTAL
1	A01	8

Execution Phase of Program 4.1

```
data sas4_2(drop=score);
     set sas4_1;
     by id;
     if first.id then total = 0;
     total + score;
     if last.id;
run;
```

SAS4_1:

	ID	SCORE
1	A01	3
2	A01	3
3	A01	2
4	A02	4
5	A02	2

PDV

_N_	D	FIRST.ID	D	LAST.ID	D	ID	K	SCORE	D	TOTAL	K
4		1		0		A02		4		0	

4th Iteration:

- FIRST.ID = 1: TOTAL ←0

SAS4_2:

	ID	TOTAL
1	A01	8

Execution Phase of Program 4.1

```
data sas4_2(drop=score);
     set sas4_1;
     by id;
     if first.id then total = 0;
     total + score;
     if last.id;
run;
```

SAS4_1:

	ID	SCORE
1	A01	3
2	A01	3
3	A01	2
4	A02	4
5	A02	2

PDV

_N_	D	FIRST.ID	D	LAST.ID	D	ID	K	SCORE	D	TOTAL	K
4		1		0		A02		4		4	

4th Iteration:

- TOTAL is calculated

SAS4_2:

	ID	TOTAL
1	A01	8

Execution Phase of Program 4.1

```
data sas4_2(drop=score);
     set sas4_1;
     by id;
     if first.id then total = 0;
     total + score;
   ➔ if last.id;
run;
```

SAS4_1:

	ID	SCORE
1	A01	3
2	A01	3
3	A01	2
4	A02	4
5	A02	2

PDV

_N_	D	FIRST.ID	D	LAST.ID	D	ID	K	SCORE	D	TOTAL	K
4		1		0		A02		4		4	

4th Iteration:

- Since LAST.ID ≠ 1(the subsetting IF statement is false), SAS immediately returns to the beginning of the DATA step

SAS4_2:

	ID	TOTAL
1	A01	8

Execution Phase of Program 4.1

```
data sas4_2(drop=score);
     set sas4_1;
     by id;
     if first.id then total = 0;
     total + score;
     if last.id;
run;
```

SAS4_1:

	ID	SCORE
1	A01	3
2	A01	3
3	A01	2
4	A02	4
5	A02	2

PDV

_N_	D	FIRST.ID	D	LAST.ID	D	ID	K	SCORE	D	TOTAL	K
5		1		0		A02		4		4	

5th Iteration:

- _N_ ↑ 5
- The values for the remaining variables are retained

SAS4_2:

	ID	TOTAL
1	A01	8

Execution Phase of Program 4.1

```
data sas4_2(drop=score);
   set sas4_1;
     by id;
     if first.id then total = 0;
     total + score;
     if last.id;
run;
```

SAS4_1:

	ID	SCORE
1	A01	3
2	A01	3
3	A01	2
4	A02	4
5	A02	2

PDV

_N_	D	FIRST.ID	D	LAST.ID	D	ID	K	SCORE	D	TOTAL	K
5		0		1		A02		2		4	

5th Iteration:

- 5th observation → PDV

SAS4_2:

	ID	TOTAL
1	A01	8

Execution Phase of Program 4.1

```
data sas4_2(drop=score);
   set sas4_1;
     by id;
     if first.id then total = 0;
     total + score;
     if last.id;
run;
```

SAS4_1:

	ID	SCORE
1	A01	3
2	A01	3
3	A01	2
4	A02	4
5	A02	2

PDV

_N_	D	FIRST.ID	D	LAST.ID	D	ID	K	SCORE	D	TOTAL	K
5		0		1		A02		2		4	

5th Iteration:

- 5th observation → PDV
- FIRST.ID ← 0
- LAST.ID ← 1

SAS4_2:

	ID	TOTAL
1	A01	8

Execution Phase of Program 4.1

```
data sas4_2(drop=score);
     set sas4_1;
     by id;
     if first.id then total = 0;
     total + score;
     if last.id;
run;
```

SAS4_1:

	ID	SCORE
1	A01	3
2	A01	3
3	A01	2
4	A02	4
5	A02	2

PDV

_N_	D	FIRST.ID	D	LAST.ID	D	ID	K	SCORE	D	TOTAL	K
5		0		1		A02		2		4	

5th Iteration:

- FIRST.ID ≠ 1: no execution

SAS4_2:

	ID	TOTAL
1	A01	8

Execution Phase of Program 4.1

```
data sas4_2(drop=score);
     set sas4_1;
     by id;
     if first.id then total = 0;
     total + score;
     if last.id;
run;
```

SAS4_1:

	ID	SCORE
1	A01	3
2	A01	3
3	A01	2
4	A02	4
5	A02	2

PDV

_N_	D	FIRST.ID	D	LAST.ID	D	ID	K	SCORE	D	TOTAL	K
5		0		1		A02		2		6	

5th Iteration:

- TOTAL is calculated

SAS4_2:

	ID	TOTAL
1	A01	8

Execution Phase of Program 4.1

```
data sas4_2(drop=score);
     set sas4_1;
     by id;
     if first.id then total = 0;
     total + score;
     if last.id;
run;
```

SAS4_1:

	ID	SCORE
1	A01	3
2	A01	3
3	A01	2
4	A02	4
5	A02	2

PDV

_N_	D	FIRST.ID	D	LAST.ID	D	ID	K	SCORE	D	TOTAL	K
5		0		1		A02		2		6	

5th Iteration:

- Since LAST.ID equals 1 (the subsetting IF statement is true), SAS continues to execute the remaining statements in the DATA step

SAS4_2:

	ID	TOTAL
1	A01	8

Execution Phase of Program 4.1

```
data sas4_2(drop=score);
     set sas4_1;
     by id;
     if first.id then total = 0;
     total + score;
     if last.id;
run;
```

SAS4_1:

	ID	SCORE
1	A01	3
2	A01	3
3	A01	2
4	A02	4
5	A02	2

PDV

_N_	D	FIRST.ID	D	LAST.ID	D	ID	K	SCORE	D	TOTAL	K
5		0		1		A02		2		6	

5th Iteration:

- The implicit OUTPUT statement copies ID and TOTAL in the PDV to the output data
- SAS returns to the beginning of the DATA step to begin the 6th iteration
- With no more observations to read in the 3rd iteration, SAS goes to the next DATA or PROC step

SAS4_2:

	ID	TOTAL
1	A01	8
2	A02	6