Select rows

1. [**How to Select the First N Rows**](https://sasexamplecode.com/how-to-select-the-first-n-rows-in-sas/)

**data** work.my\_ds;

do i="01JAN2020"d to "10JAN2020"d;

date = i;

weekday = compress(put(i, dowName.));

output;

end;

drop i;

format date date9.;

**run**;

**proc sql** outobs=**5**;

create table work.first\_5\_obs\_sql as

select *\* from work.my\_ds;*

**quit**;

**proc sql** inobs=**5**;

create table work.first\_5\_obs\_sql as

select *\* from work.my\_ds;*

**quit**;

## Select the First N Rows

### Method II is more efficient than Method I. Unlike Method I, SAS stops processing the data once the N-th observation from the input data set is read. The Method I, with the \_N\_ variable, processes all rows from the input data set and checks for each row if the IF-statement holds.

### Method I

**data** work.first\_5\_obs\_sas;

set work.my\_ds;

if \_N\_ le **5** then output;

**run**;

### Method II

### Another method to select the first N rows from a dataset is using the [***OBS=***-option](https://documentation.sas.com/?cdcId=pgmsascdc&cdcVersion=9.4_3.5&docsetId=ledsoptsref&docsetTarget=p0h5nwbig8mobbn1u0dwtdo0c0a0.htm&locale=en). With this option, you ca Method II

n specify the last row that SAS processes from the input datase

**data** work.first\_5\_obs\_sas;

set work.my\_ds (obs=**5**);

**run**;

## Select N Observations Randomly

/\*

SRS method (simple random sampling)

\*/

**proc surveyselect** **data**=work.my\_ds

out=work.random\_obs

method=srs

sampsize=**4**

seed=**123456789**;

**run**;

[How to Select the First Row of a Group in SAS](https://sasexamplecode.com/how-to-select-the-first-row-of-a-group-in-sas/)

**data** work.my\_ds;

infile datalines dlm=",";

length name race $25. race\_result **8**;

format race\_date date9.;

input name $ race $ race\_result race\_date :date9.;

datalines;

John, Boston Marathon, **7**, 15APR2019

John, NY Marathon, **5**, 4NOV2018

John, NY Marathon, **12**, 3NOV2019

Maria, Boston Marathon, **8**, 15APR2019

Maria, Chicago Marathon, **4**, 13OCT2019

Will, NY Marathon, **25**, 4NOV2018

Emma, Chicago Marathon, **11**, 7OCT2018

Emma, Chicago Marathon, **9**, 13OCT2019

Juan, NY Marathon, **3**, 3NOV2019

;

**run**;

**proc sort** **data**=work.my\_ds

out=work.my\_ds\_srt;

by race\_result;

**run**;

**data** work.best\_race\_result;

set work.my\_ds\_srt(obs=**1**);

**run**;

**data** work.worst\_race\_result;

set work.my\_ds\_srt end=last\_row;

if last\_row then output;

**run**

**proc sort** **data**=work.my\_ds

out=work.my\_ds\_srt;

by race race\_result;

**run**;

**data** work.first\_by\_group;

set work.my\_ds\_srt;

by race;

if first.race then output;

**run**;

**data** work.last\_by\_group;

set work.my\_ds\_srt;

by race;

if last.race then output;

**run**;

**proc sql**;

create table work.first\_by\_group\_sql as

select name,

race,

race\_result,

race\_date

from work.my\_ds

group by race

having min(race\_result) eq race\_result;

**quit**;

**data** work.first\_2\_by\_group;

set work.my\_ds\_srt;

by race;

retain n;

if first.race then do;

n = **1**;

output;

end;

else if n lt **2** then do;

n = n + **1**;

output;

end;

**run**;

## Select the First N Rows by Group with PROC RANK

proc rank **data**=work.my\_ds\_srt

out=work.my\_ds\_ranked;

var race\_result;

by race;

ranks group\_rank;

**run**;

**data** work.first\_2\_by\_group;

set work.my\_ds\_ranked;

if group\_rank le **2** then output;

**run**;