calculate-the-average-per-group

<https://sasexamplecode.com/5-ways-in-sas-to-calculate-the-average-per-group/>

1. USE SQL

data work.exam\_results;

input student $ exam $ result;

datalines;

Jack Physics 90

Jack Biology 80

Jack Geometry 92

Will Geometry 70

Will Biology 72

Maria Physics 85

Emma Biology 90

Emma Physics 75

Jim Biology 65

Jim Physics 73

;

run;

proc print data=exam\_results noobs;

run;

proc sql;

create table work.avg\_by\_group\_1 as

select student,

avg(result) as avg\_result format=comma12.2

from work.exam\_results

group by student

order by calculated avg\_result desc;

quit;

/\* The FEEDBACK option displays the columns that are represented by a SELECT \* statement.

* The STIMER option records and displays query execution time. T
* he NUMBER option displays the row number in the result table.
* AVG() function is the same as MEAN() function. \*/

proc sql feedback number stimer;

select student,

avg(result) as avg\_result format=comma12.2

from work.exam\_results

group by student

order by avg\_result desc;

quit;

| Row | student | avg\_result |
| --- | --- | --- |
| 1 | Jack | 87.33 |
| 2 | Maria | 85.00 |
| 3 | Emma | 82.50 |
| 4 | Will | 71.00 |
| 5 | Jim | 69.00 |

1. Method 2: PROC MEANS

/\* group by student \*/

proc means data=work.exam\_results nway;

class student;

var result;

run;

The MEANS Procedure

| Analysis Variable : result | | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
| student | N Obs | N | Mean | Std Dev | Minimum | Maximum |
| Emma | 2 | 2 | 82.5000000 | 10.6066017 | 75.0000000 | 90.0000000 |
| Jack | 3 | 3 | 87.3333333 | 6.4291005 | 80.0000000 | 92.0000000 |
| Jim | 2 | 2 | 69.0000000 | 5.6568542 | 65.0000000 | 73.0000000 |
| Maria | 1 | 1 | 85.0000000 | . | 85.0000000 | 85.0000000 |
| Will | 2 | 2 | 71.0000000 | 1.4142136 | 70.0000000 | 72.0000000 |

/\*get mean and Assigned name 'avg\_result' \*/

proc means data=work.exam\_results nway;

class student;

var result;

output out=work.avg\_by\_group\_2 (drop = \_TYPE\_ \_FREQ\_)

mean=avg\_result;

run;

proc contents data = work.avg\_by\_group\_2; run;

proc print data = work.avg\_by\_group\_2 noobs; run;

1. Use data step

proc sort data=work.exam\_results out=work.exam\_results\_srt; by student; run;

data work.avg\_by\_group\_5 ;

set work.exam\_results\_srt;

by student;

run;

| student | exam | result |
| --- | --- | --- |
| Emma | Biology | 90 |
| Emma | Physics | 75 |
| Jack | Physics | 90 |
| Jack | Biology | 80 |
| Jack | Geometry | 92 |
| Jim | Biology | 65 |
| Jim | Physics | 73 |
| Maria | Physics | 85 |
| Will | Geometry | 70 |
| Will | Biology | 72 |

/\*Use data step \*/

proc sort data=work.exam\_results out=work.exam\_results\_srt; by student; run;

data work.avg\_by\_group\_5 ;

set work.exam\_results\_srt;

by student;

run;

proc print data = work.avg\_by\_group\_5 noobs; run;

proc sort data=work.exam\_results

out=work.exam\_results\_srt;

by student;

run;

data work.avg\_by\_group\_5 ;

set work.exam\_results\_srt;

by student;

retain n\_exams sum\_results avg\_result;

if first.student then do;

n\_exams = 1;

sum\_results = result;

avg\_result = result;

end;

else do;

n\_exams = n\_exams + 1;

sum\_results = sum\_results + result;

avg\_result = sum\_results / n\_exams;

end;

if last.student then output;

drop exam result;

format avg\_result comma12.2;

run;

proc print data = work.avg\_by\_group\_5 noobs; run;

| **student** | **n\_exams** | **sum\_results** | **avg\_result** |
| --- | --- | --- | --- |
| Emma | 2 | 165 | 82.50 |
| Jack | 3 | 262 | 87.33 |
| Jim | 2 | 138 | 69.00 |
| Maria | 1 | 85 | 85.00 |
| Will | 2 | 142 | 71.00 |