

Using Multiple DOW Loops

Dr. Arthur Tabachneck
Director, Data Management

Idea stolen from a SAS Global Forum paper by
Paul Dorfman and Koen Vyverman

<http://support.sas.com/resources/papers/proceedings09/038-2009.pdf>

suppose you had the following data:

	Name	Sex	Age	Height	Weight
1	Alice	F	13	56.5	84
2	Barbara	F	13	65.3	98
3	Carol	F	14	62.8	102.5
4	Jane	F	12	59.8	84.5
5	Janet	F	15	62.5	112.5
6	Joyce	F	11	51.3	50.5
7	Judy	F	14	64.3	90
8	Louise	F	12	56.3	77
9	Mary	F	15	66.5	112
10	Alfred	M	14	69	112.5
11	Henry	M	14	63.5	102.5
12	James	M	12	57.3	83
13	Jeffrey	M	13	62.5	84
14	John	M	12	59	99.5
15	Philip	M	16	72	150
16	Robert	M	12	64.8	128
17	Ronald	M	15	67	133
18	Thomas	M	11	57.5	85
19	William	M	15	66.5	112

i.e.,
 proc sort data=sashelp.class out=have;
 by sex;
 run;



and you needed to have the following table:

	Name	Sex	Age	Height	Weight	z
1	Alice	F	13	56.5	84	-0.814791157
2	Barbara	F	13	65.3	98	0.938781116
3	Carol	F	14	62.8	102.5	0.4406071747
4	Jane	F	12	59.8	84.5	-0.157201555
5	Janet	F	15	62.5	112.5	0.3808263018
6	Joyce	F	11	51.3	50.5	-1.850992955
7	Judy	F	14	64.3	90	0.7395115395
8	Louise	F	12	56.3	77	-0.854645073
9	Mary	F	15	66.5	112	1.1779046079
10	Alfred	M	14	69	112.5	1.0307948355
11	Henry	M	14	63.5	102.5	-0.083030625
12	James	M	12	57.3	83	-1.33861569
13	Jeffrey	M	13	62.5	84	-0.285544345
14	John	M	12	59	99.5	-0.994342366
15	Philip	M	16	72	150	1.6383359959
16	Robert	M	12	64.8	128	0.1802372109
17	Ronald	M	15	67	133	0.6257673952
18	Thomas	M	11	57.5	85	-1.298112946
19	William	M	15	66.5	112	0.5245105352

i.e, with a new variable that shows height in terms of number of standard deviations from the group average

separately, for males and females:

$(\text{height} - \text{average}) / \text{sd}$

Note: this presentation is about the method .. not the particular calculation. Of course, for some calculations, certain procs might be more advantageous



did you know you can do it in a single data step?

```
data want (drop=n);
do until(last.sex);
  set have;
  by sex;
  avg=sum(avg,height);
  n=sum(n,height/height);
end;
avg=avg/n;
do until(last.sex);
  set have;
  by sex;
  sd=sum(sd,(height-avg)**2);
end;
sd=sqrt(sd/(n-1));
do until(last.sex);
  set have;
  by sex;
  if missing(height) then call missing(z);
  else z=(height-avg)/sd;
  output;
end;
run;
```



Why it works: Look under the hood:

PDV →

Last.sex	Name	Sex	Age	Height	Weight	First.Sex	avg	n	sd	z	_ERROR_	_N_
1	1	0	1

```
put _all_; → data want (drop=n);
               do until(last.sex);
                 set have;
                 by sex;
                 avg=sum(avg,height);
                 n=sum(n,height/height);
               end;
               avg=avg/n;
               do until(last.sex);
                 set have;
                 by sex;
                 sd=sum(sd,(height-avg)**2);
               end;
               sd=sqrt(sd/(n-1));
               do until(last.sex);
                 set have;
                 by sex;
                 if missing(height) then call missing(z);
                 else z=(height-avg)/sd;
                 output;
               end;
run;
```



Why it works: Under the hood-Loop #1:

PDV →

Last.sex	Name	Sex	Age	Height	Weight	First.Sex	avg	n	sd	z	_ERROR_	_N_
0	Alice	F	13	56.5	84	1	56.5	1	.	.	0	1

```
data want (drop=n);  
do until(last.sex);  
  set have;  
  by sex;  
  avg=sum(avg,height);  
  n=sum(n,height/height);  
end;  
put _all_; →  
avg=avg/n;  
do until(last.sex);  
  set have;  
  by sex;  
  sd=sum(sd,(height-avg)**2);  
end;  
sd=sqrt(sd/(n-1));  
do until(last.sex);  
  set have;  
  by sex;  
  if missing(height) then call missing(z);  
  else z=(height-avg)/sd;  
  output;  
end;  
run;
```



Why it works: Under the hood-Loop #1:

PDV →

Last.sex	Name	Sex	Age	Height	Weight	First.Sex	avg	n	sd	z	_ERROR_	_N_
0	Barbara	F	13	65.3	98	0	121.8	2	.	.	0	1

```

data want (drop=n);
do until(last.sex);
  set have;
  by sex;
  avg=sum(avg,height);
  n=sum(n,height/height);
end;
put _all_; →
avg=avg/n;
do until(last.sex);
  set have;
  by sex;
  sd=sum(sd,(height-avg)**2);
end;
sd=sqrt(sd/(n-1));
do until(last.sex);
  set have;
  by sex;
  if missing(height) then call missing(z);
  else z=(height-avg)/sd;
  output;
end;
run;

```



Why it works: Under the hood-Loop #1:

PDV →

Last.sex	Name	Sex	Age	Height	Weight	First.Sex	avg	n	sd	z	_ERROR_	_N_
0	Carol	F	14	62.8	102.5	0	184.6	3	.	.	0	1

```

data want (drop=n);
do until(last.sex);
  set have;
  by sex;
  avg=sum(avg,height);
  n=sum(n,height/height);
end;
put _all_; →
avg=avg/n;
do until(last.sex);
  set have;
  by sex;
  sd=sum(sd,(height-avg)**2);
end;
sd=sqrt(sd/(n-1));
do until(last.sex);
  set have;
  by sex;
  if missing(height) then call missing(z);
  else z=(height-avg)/sd;
  output;
end;
run;

```



Why it works: Under the hood-Loop #1:

PDV →

Last.sex	Name	Sex	Age	Height	Weight	First.Sex	avg	n	sd	z	_ERROR_	_N_
0	Jane	F	12	59.8	84.5	0	244.4	4	.	.	0	1

```
data want (drop=n);  
do until(last.sex);  
  set have;  
  by sex;  
  avg=sum(avg,height);  
  n=sum(n,height/height);  
end;  
put _all_; →  
avg=avg/n;  
do until(last.sex);  
  set have;  
  by sex;  
  sd=sum(sd,(height-avg)**2);  
end;  
sd=sqrt(sd/(n-1));  
do until(last.sex);  
  set have;  
  by sex;  
  if missing(height) then call missing(z);  
  else z=(height-avg)/sd;  
  output;  
end;  
run;
```



Why it works: Under the hood-Loop #1:

PDV →

Last.sex	Name	Sex	Age	Height	Weight	First.Sex	avg	n	sd	z	_ERROR_	_N_
0	Janet	F	15	62.5	112.5	0	306.9	5	.	.	0	1

```
data want (drop=n);  
do until(last.sex);  
  set have;  
  by sex;  
  avg=sum(avg,height);  
  n=sum(n,height/height);  
end;  
put _all_; →  
avg=avg/n;  
do until(last.sex);  
  set have;  
  by sex;  
  sd=sum(sd,(height-avg)**2);  
end;  
sd=sqrt(sd/(n-1));  
do until(last.sex);  
  set have;  
  by sex;  
  if missing(height) then call missing(z);  
  else z=(height-avg)/sd;  
  output;  
end;  
run;
```



Why it works: Under the hood-Loop #1:

PDV →

Last.sex	Name	Sex	Age	Height	Weight	First.Sex	avg	n	sd	z	_ERROR_	_N_
0	Joyce	F	11	51.3	50.5	0	358.2	6	.	.	0	1

```
data want (drop=n);
do until(last.sex);
  set have;
  by sex;
  avg=sum(avg,height);
  n=sum(n,height/height);
end;
put _all_; →
avg=avg/n;
do until(last.sex);
  set have;
  by sex;
  sd=sum(sd,(height-avg)**2);
end;
sd=sqrt(sd/(n-1));
do until(last.sex);
  set have;
  by sex;
  if missing(height) then call missing(z);
  else z=(height-avg)/sd;
  output;
end;
run;
```

Why it works: Under the hood-Loop #1:

PDV →

Last.sex	Name	Sex	Age	Height	Weight	First.Sex	avg	n	sd	z	_ERROR_	_N_
0	Judy	F	14	64.3	90	0	422.5	7	.	.	0	1

```

data want (drop=n);
do until(last.sex);
  set have;
  by sex;
  avg=sum(avg,height);
  n=sum(n,height/height);
end;
put _all_; →
avg=avg/n;
do until(last.sex);
  set have;
  by sex;
  sd=sum(sd,(height-avg)**2);
end;
sd=sqrt(sd/(n-1));
do until(last.sex);
  set have;
  by sex;
  if missing(height) then call missing(z);
  else z=(height-avg)/sd;
  output;
end;
run;
    
```



Why it works: Under the hood-Loop #1:

PDV →

Last.sex	Name	Sex	Age	Height	Weight	First.Sex	avg	n	sd	z	_ERROR_	_N_
0	Louise	F	12	56.3	77	0	478.8	8	.	.	0	1

```
data want (drop=n);  
do until(last.sex);  
  set have;  
  by sex;  
  avg=sum(avg,height);  
  n=sum(n,height/height);  
end;  
put _all_; →  
avg=avg/n;  
do until(last.sex);  
  set have;  
  by sex;  
  sd=sum(sd,(height-avg)**2);  
end;  
sd=sqrt(sd/(n-1));  
do until(last.sex);  
  set have;  
  by sex;  
  if missing(height) then call missing(z);  
  else z=(height-avg)/sd;  
  output;  
end;  
run;
```



Why it works: Under the hood-Loop #1:

PDV →

Last.sex	Name	Sex	Age	Height	Weight	First.Sex	avg	n	sd	z	_ERROR_	_N_
1	Mary	F	15	66.5	112	0	545.3	9	.	.	0	1

```
data want (drop=n);  
do until(last.sex);  
  set have;  
  by sex;  
  avg=sum(avg,height);  
  n=sum(n,height/height);  
end;  
put _all_; →  
avg=avg/n;  
do until(last.sex);  
  set have;  
  by sex;  
  sd=sum(sd,(height-avg)**2);  
end;  
sd=sqrt(sd/(n-1));  
do until(last.sex);  
  set have;  
  by sex;  
  if missing(height) then call missing(z);  
  else z=(height-avg)/sd;  
  output;  
end;  
run;
```



Why it works: Under the hood-Loop #1:

PDV →

Last.sex	Name	Sex	Age	Height	Weight	First.Sex	avg	n	sd	z	_ERROR_	_N_
1	Mary	F	15	66.5	112	0	60.59	9	.	.	0	1

```
data want (drop=n);
do until(last.sex);
  set have;
  by sex;
  avg=sum(avg,height);
  n=sum(n,height/height);
end;
put _all_; → avg=avg/n;
do until(last.sex);
  set have;
  by sex;
  sd=sum(sd,(height-avg)**2);
end;
sd=sqrt(sd/(n-1));
do until(last.sex);
  set have;
  by sex;
  if missing(height) then call missing(z);
  else z=(height-avg)/sd;
  output;
end;
run;
```



Why it works: Under the hood-Loop #2:

PDV →

Last.sex	Name	Sex	Age	Height	Weight	First.Sex	avg	n	sd	z	_ERROR_	_N_
0	Alice	F	13	56.5	84	1	60.59	9	16.719	.	0	1

```

data want (drop=n);
do until(last.sex);
  set have;
  by sex;
  avg=sum(avg,height);
  n=sum(n,height/height);
end;
avg=avg/n;
do until(last.sex);
  set have;
  by sex;
  sd=sum(sd,(height-avg)**2);
end;
put _all_; →
sd=sqrt(sd/(n-1));
do until(last.sex);
  set have;
  by sex;
  if missing(height) then call missing(z);
  else z=(height-avg)/sd;
  output;
end;
run;

```



Why it works: Under the hood-Loop #2:

PDV →

Last.sex	Name	Sex	Age	Height	Weight	First.Sex	avg	n	sd	z	_ERROR_	_N_
0	Barbara	F	13	65.3	98	0	60.59	9	38.914	.	0	1

```
data want (drop=n);
do until(last.sex);
  set have;
  by sex;
  avg=sum(avg,height);
  n=sum(n,height/height);
end;
avg=avg/n;
do until(last.sex);
  set have;
  by sex;
  sd=sum(sd,(height-avg)**2);
end;
put _all_; →
sd=sqrt(sd/(n-1));
do until(last.sex);
  set have;
  by sex;
  if missing(height) then call missing(z);
  else z=(height-avg)/sd;
  output;
end;
run;
```



Why it works: Under the hood-Loop #2:

PDV →

Last.sex	Name	Sex	Age	Height	Weight	First.Sex	avg	n	sd	z	_ERROR_	_N_
0	Carol	F	14	62.8	102.5	0	60.59	9	43.803	.	0	1

```

data want (drop=n);
do until(last.sex);
  set have;
  by sex;
  avg=sum(avg,height);
  n=sum(n,height/height);
end;
avg=avg/n;
do until(last.sex);
  set have;
  by sex;
  sd=sum(sd,(height-avg)**2);
end;
put _all_; →
sd=sqrt(sd/(n-1));
do until(last.sex);
  set have;
  by sex;
  if missing(height) then call missing(z);
  else z=(height-avg)/sd;
  output;
end;
run;

```



Why it works: Under the hood-Loop #2:

PDV →

Last.sex	Name	Sex	Age	Height	Weight	First.Sex	avg	n	sd	z	_ERROR_	_N_
0	Jane	F	12	59.8	84.5	0	60.59	9	44.425	.	0	1

```
data want (drop=n);
do until(last.sex);
  set have;
  by sex;
  avg=sum(avg,height);
  n=sum(n,height/height);
end;
avg=avg/n;
do until(last.sex);
  set have;
  by sex;
  sd=sum(sd,(height-avg)**2);
end;
put _all_; →
sd=sqrt(sd/(n-1));
do until(last.sex);
  set have;
  by sex;
  if missing(height) then call missing(z);
  else z=(height-avg)/sd;
  output;
end;
run;
```

Why it works: Under the hood-Loop #2:

PDV →

Last.sex	Name	Sex	Age	Height	Weight	First.Sex	avg	n	sd	z	_ERROR_	_N_
0	Janet	F	15	62.5	112.5	0	60.59	9	48.077	.	0	1

```

data want (drop=n);
do until(last.sex);
  set have;
  by sex;
  avg=sum(avg,height);
  n=sum(n,height/height);
end;
avg=avg/n;
do until(last.sex);
  set have;
  by sex;
  sd=sum(sd,(height-avg)**2);
end;
sd=sqrt(sd/(n-1));
do until(last.sex);
  set have;
  by sex;
  if missing(height) then call missing(z);
  else z=(height-avg)/sd;
  output;
end;
run;
    
```

put _all_; →



Why it works: Under the hood-Loop #2:

PDV →

Last.sex	Name	Sex	Age	Height	Weight	First.Sex	avg	n	sd	z	_ERROR_	_N_
0	Joyce	F	11	51.3	50.5	0	60.59	9	134.361	.	0	1

```
data want (drop=n);
do until(last.sex);
  set have;
  by sex;
  avg=sum(avg,height);
  n=sum(n,height/height);
end;
avg=avg/n;
do until(last.sex);
  set have;
  by sex;
  sd=sum(sd,(height-avg)**2);
end;
put _all_; →
sd=sqrt(sd/(n-1));
do until(last.sex);
  set have;
  by sex;
  if missing(height) then call missing(z);
  else z=(height-avg)/sd;
  output;
end;
run;
```



Why it works: Under the hood-Loop #2:

PDV →

Last.sex	Name	Sex	Age	Height	Weight	First.Sex	avg	n	sd	z	_ERROR_	_N_
0	Judy	F	14	64.3	90	0	60.59	9	148.133	.	0	1

```
data want (drop=n);
do until(last.sex);
  set have;
  by sex;
  avg=sum(avg,height);
  n=sum(n,height/height);
end;
avg=avg/n;
do until(last.sex);
  set have;
  by sex;
  sd=sum(sd,(height-avg)**2);
end;
put _all_; →
sd=sqrt(sd/(n-1));
do until(last.sex);
  set have;
  by sex;
  if missing(height) then call missing(z);
  else z=(height-avg)/sd;
  output;
end;
run;
```



Why it works: Under the hood-Loop #2:

PDV →

Last.sex	Name	Sex	Age	Height	Weight	First.Sex	avg	n	sd	z	_ERROR_	_N_
0	Louise	F	12	56.3	77	0	60.59	9	166.528	.	0	1

```
data want (drop=n);
do until(last.sex);
  set have;
  by sex;
  avg=sum(avg,height);
  n=sum(n,height/height);
end;
avg=avg/n;
do until(last.sex);
  set have;
  by sex;
  sd=sum(sd,(height-avg)**2);
end;
put _all_; →
sd=sqrt(sd/(n-1));
do until(last.sex);
  set have;
  by sex;
  if missing(height) then call missing(z);
  else z=(height-avg)/sd;
  output;
end;
run;
```

Why it works: Under the hood-Loop #2:

PDV →

Last.sex	Name	Sex	Age	Height	Weight	First.Sex	avg	n	sd	z	_ERROR_	_N_
1	Mary	F	15	66.5	112	0	60.59	9	201.469	.	0	1

```

data want (drop=n);
do until(last.sex);
  set have;
  by sex;
  avg=sum(avg,height);
  n=sum(n,height/height);
end;
avg=avg/n;
do until(last.sex);
  set have;
  by sex;
  sd=sum(sd,(height-avg)**2);
end;
sd=sqrt(sd/(n-1));
do until(last.sex);
  set have;
  by sex;
  if missing(height) then call missing(z);
  else z=(height-avg)/sd;
  output;
end;
run;
    
```

put _all_; →



Why it works: Under the hood-Loop #2:

PDV →

Last.sex	Name	Sex	Age	Height	Weight	First.Sex	avg	n	sd	z	_ERROR_	_N_
1	Mary	F	15	66.5	112	0	60.59	9	5.018	.	0	1

```

data want (drop=n);
do until(last.sex);
  set have;
  by sex;
  avg=sum(avg,height);
  n=sum(n,height/height);
end;
avg=avg/n;
do until(last.sex);
  set have;
  by sex;
  sd=sum(sd,(height-avg)**2);
end;
put _all_; → sd=sqrt(sd/(n-1));
do until(last.sex);
  set have;
  by sex;
  if missing(height) then call missing(z);
  else z=(height-avg)/sd;
  output;
end;
run;

```



Why it works: Under the hood-Loop #3:

PDV →

Last.sex	Name	Sex	Age	Height	Weight	First.Sex	avg	n	sd	z	_ERROR_	_N_
0	Alice	F	13	56.5	84	1	60.59	9	5.018	-.81	0	1

```

data want (drop=n);
do until(last.sex);
  set have;
  by sex;
  avg=sum(avg,height);
  n=sum(n,height/height);
end;
avg=avg/n;
do until(last.sex);
  set have;
  by sex;
  sd=sum(sd,(height-avg)**2);
end;
sd=sqrt(sd/(n-1));
do until(last.sex);
  set have;
  by sex;
  if missing(height) then call missing(z);
  else z=(height-avg)/sd;
  output;
end;
run;

```

put _all_; →



Why it works: Under the hood-Loop #3:

PDV →

Last.sex	Name	Sex	Age	Height	Weight	First.Sex	avg	n	sd	z	_ERROR_	_N_
0	Barbara	F	13	65.3	98	0	60.59	9	5.018	-.94	0	1

```
data want (drop=n);
do until(last.sex);
  set have;
  by sex;
  avg=sum(avg,height);
  n=sum(n,height/height);
end;
avg=avg/n;
do until(last.sex);
  set have;
  by sex;
  sd=sum(sd,(height-avg)**2);
end;
sd=sqrt(sd/(n-1));
do until(last.sex);
  set have;
  by sex;
  if missing(height) then call missing(z);
  else z=(height-avg)/sd;
  output;
end;
put _all_; →
run;
```

Why it works: Under the hood-Loop #3:

PDV →

Last.sex	Name	Sex	Age	Height	Weight	First.Sex	avg	n	sd	z	_ERROR_	_N_
0	Carol	F	14	62.8	102.5	0	60.59	9	5.018	.44	0	1

```

data want (drop=n);
do until(last.sex);
  set have;
  by sex;
  avg=sum(avg,height);
  n=sum(n,height/height);
end;
avg=avg/n;
do until(last.sex);
  set have;
  by sex;
  sd=sum(sd,(height-avg)**2);
end;
sd=sqrt(sd/(n-1));
do until(last.sex);
  set have;
  by sex;
  if missing(height) then call missing(z);
  else z=(height-avg)/sd;
  output;
end;
put _all_; →
run;
  
```



Why it works: Under the hood-Loop #3:

PDV →

Last.sex	Name	Sex	Age	Height	Weight	First.Sex	avg	n	sd	z	_ERROR_	_N_
0	Jane	F	12	59.8	84.5	0	60.59	9	5.018	-.16	0	1

```

data want (drop=n);
do until(last.sex);
  set have;
  by sex;
  avg=sum(avg,height);
  n=sum(n,height/height);
end;
avg=avg/n;
do until(last.sex);
  set have;
  by sex;
  sd=sum(sd,(height-avg)**2);
end;
sd=sqrt(sd/(n-1));
do until(last.sex);
  set have;
  by sex;
  if missing(height) then call missing(z);
  else z=(height-avg)/sd;
  output;
end;
put _all_; →
run;

```

Why it works: Under the hood-Loop #3:

PDV →

Last.sex	Name	Sex	Age	Height	Weight	First.Sex	avg	n	sd	z	_ERROR_	_N_
0	Janet	F	15	62.5	112.5	0	60.59	9	5.018	.38	0	1

```

data want (drop=n);
do until(last.sex);
  set have;
  by sex;
  avg=sum(avg,height);
  n=sum(n,height/height);
end;
avg=avg/n;
do until(last.sex);
  set have;
  by sex;
  sd=sum(sd,(height-avg)**2);
end;
sd=sqrt(sd/(n-1));
do until(last.sex);
  set have;
  by sex;
  if missing(height) then call missing(z);
  else z=(height-avg)/sd;
  output;
end;
put _all_; →
run;
    
```



Why it works: Under the hood-Loop #3:

PDV →

Last.sex	Name	Sex	Age	Height	Weight	First.Sex	avg	n	sd	z	_ERROR_	_N_
0	Judy	F	14	64.3	90	0	60.59	9	5.018	.74	0	1

```
data want (drop=n);
do until(last.sex);
  set have;
  by sex;
  avg=sum(avg,height);
  n=sum(n,height/height);
end;
avg=avg/n;
do until(last.sex);
  set have;
  by sex;
  sd=sum(sd,(height-avg)**2);
end;
sd=sqrt(sd/(n-1));
do until(last.sex);
  set have;
  by sex;
  if missing(height) then call missing(z);
  else z=(height-avg)/sd;
  output;
end;
put _all_; →
run;
```



Why it works: Under the hood-Loop #3:

PDV →

Last.sex	Name	Sex	Age	Height	Weight	First.Sex	avg	n	sd	z	_ERROR_	_N_
0	Louise	F	12	56.3	77	0	60.59	9	5.018	-.85	0	1

```
data want (drop=n);
do until(last.sex);
  set have;
  by sex;
  avg=sum(avg,height);
  n=sum(n,height/height);
end;
avg=avg/n;
do until(last.sex);
  set have;
  by sex;
  sd=sum(sd,(height-avg)**2);
end;
sd=sqrt(sd/(n-1));
do until(last.sex);
  set have;
  by sex;
  if missing(height) then call missing(z);
  else z=(height-avg)/sd;
  output;
end;
put _all_; →
run;
```



Why it works: Under the hood-Loop #3:

PDV →

Last.sex	Name	Sex	Age	Height	Weight	First.Sex	avg	n	sd	z	_ERROR_	_N_
1	Mary	F	15	66.5	112	0	60.59	9	5.018	1.18	0	1

```
data want (drop=n);
do until(last.sex);
  set have;
  by sex;
  avg=sum(avg,height);
  n=sum(n,height/height);
end;
avg=avg/n;
do until(last.sex);
  set have;
  by sex;
  sd=sum(sd,(height-avg)**2);
end;
sd=sqrt(sd/(n-1));
do until(last.sex);
  set have;
  by sex;
  if missing(height) then call missing(z);
  else z=(height-avg)/sd;
  output;
end;
put _all_; →
run;
```



Why it works: Under the hood-2nd Iteration-Loop #1:

PDV →

Last.sex	Name	Sex	Age	Height	Weight	First.Sex	avg	n	sd	z	_ERROR_	_N_
1	Mary	F	15	66.5	112	0	0	2

```

put _all_; → data want (drop=n);
              do until(last.sex);
                set have;
                by sex;
                avg=sum(avg,height);
                n=sum(n,height/height);
              end;
              avg=avg/n;
              do until(last.sex);
                set have;
                by sex;
                sd=sum(sd,(height-avg)**2);
              end;
              sd=sqrt(sd/(n-1));
              do until(last.sex);
                set have;
                by sex;
                if missing(height) then call missing(z);
                else z=(height-avg)/sd;
                output;
              end;
run;

```



Why it works: Under the hood-2nd Iteration-Loop #1:

PDV →

Last.sex	Name	Sex	Age	Height	Weight	First.Sex	avg	n	sd	z	_ERROR_	_N_
0	Alfred	M	14	69	112.5	1	69	1	.	.	0	2

```
data want (drop=n);
do until(last.sex);
  set have;
  by sex;
  avg=sum(avg,height);
  n=sum(n,height/height);
end;
put _all_; →
avg=avg/n;
do until(last.sex);
  set have;
  by sex;
  sd=sum(sd,(height-avg)**2);
end;
sd=sqrt(sd/(n-1));
do until(last.sex);
  set have;
  by sex;
  if missing(height) then call missing(z);
  else z=(height-avg)/sd;
  output;
end;
run;
```



and, after completing all iterations, of course:

	Name	Sex	Age	Height	Weight	z
1	Alice	F	13	56.5	84	-0.814791157
2	Barbara	F	13	65.3	98	0.938781116
3	Carol	F	14	62.8	102.5	0.4406071747
4	Jane	F	12	59.8	84.5	-0.157201555
5	Janet	F	15	62.5	112.5	0.3808263018
6	Joyce	F	11	51.3	50.5	-1.850992955
7	Judy	F	14	64.3	90	0.7395115395
8	Louise	F	12	56.3	77	-0.854645073
9	Mary	F	15	66.5	112	1.1779046079
10	Alfred	M	14	69	112.5	1.0307948355
11	Henry	M	14	63.5	102.5	-0.083030625
12	James	M	12	57.3	83	-1.33861569
13	Jeffrey	M	13	62.5	84	-0.285544345
14	John	M	12	59	99.5	-0.994342366
15	Philip	M	16	72	150	1.6383359959
16	Robert	M	12	64.8	128	0.1802372109
17	Ronald	M	15	67	133	0.6257673952
18	Thomas	M	11	57.5	85	-1.298112946
19	William	M	15	66.5	112	0.5245105352



Questions?

Your comments and questions are valued and encouraged.

Contact the author:

Dr. Arthur Tabachneck
Director, Data Management
Insurance Bureau of Canada
Toronto, Ontario L3T 5K9
Email: atabachneck@ibc.ca