

1 Creating a simulated dataset in SAS

The following code will create a simulated dataset with 1000 observations drawn from a normal distribution with mean $\mu = 2$ and standard deviation $\sigma = 1$.

```
options linesize = 72 ;

data symm ;
seed = 32542 ;
do i = 1 to 1000 ;
    y = 2 + rannor(seed) ;
    output ;
end ;
drop seed ;
run ;
```

We are going to treat this simulated dataset as a population – this is the complete set of items in which we are interested. To display just the first 10 records of this dataset,

```
proc print data = symm (obs=10) ;
run ;
```

2 Proc means

We can use *proc means* to get various summary statistics in a more compact format than *proc univariate* provides. The default statistics provided are

- n = number of observations
- mean
- std dev = standard deviation
- minimum
- maximum

```
proc means data = symm ;
var y ;
run ;
```

Output:

Analysis Variable : Y

N	Mean	Std Dev	Minimum	Maximum
1000	2.0142510	0.9967128	-1.3402900	5.3321347

3 Drawing simple random samples from our population

We will use *proc plan* to draw a simple random sample of size 10 from our “population” of 1000 values. We can then use *proc means* to get summary statistics for our simple random sample by copying the selected record id’s from the *proc plan* output into the code for *proc means*.

```
proc plan seed = 25;
factors a =10 of 1000 ;
run ;

proc means data = symm ;
var y ;
where i in(625 253 996 487 44 236 578 298 251 668) ;
run ;
```

N	Mean	Std Dev	Minimum	Maximum
10	2.1224623	1.1918346	-0.0975871	3.9326700

4 Drawing several different samples, and record sample statistics (mean and standard deviation) of each

1. Do the following:
 - (a) Draw a simple random sample of size 10 from our simulated “population.” Use a different seed each time so you get different samples.
 - (b) Calculate the sample mean and the sample standard deviation from the sample and record them in the table in the end of this lab worksheet.

2. Sample means from a skewed distribution.

- (a) Use the code below to simulate a dataset from a skewed distribution.

```
data skewed ;
seed = 325 ;
do i = 1 to 1000 ;
y = 2 + rangam(seed, 2) ;
output ;
end ;
drop seed ;
run ;
```

- (b) Use *proc univariate* to verify that you got a skewed distribution.

```
proc univariate plot data = skewed ;
var y ;
run ;
```

- (c) Draw a simple random sample of size 10 from the skewed population. Calculate the sample mean and the sample standard deviation from the sample and record them in the table in the end of this lab worksheet.

Symmetric, n=10	$\bar{x}$	s
Skewed, n=10	$\bar{x}$	s