

## Description

The data analyzed in [SEM] [Example 20](#) are summary statistics data (SSD) and contain summary statistics on two groups of subjects, those from grade 4 and those from grade 5. Below we show how we created this summary statistics dataset.

See [SEM] [Intro 11](#) for background on SSD.

## Remarks and examples

See [SEM] [Example 2](#) for creating a single-group dataset from published covariances. In this example, we will create a two-group dataset from published correlations, standard deviations, and means.

[Marsh and Hocevar \(1985\)](#) publish lots of SSD, of which we will enter the data for students in grade 4 and grade 5 found on pages 579–581. In that source, the authors published the correlations, standard deviations, and means of their variables.

We will set the data for the first group, declare that we have groups and wish to add another, and set the data for the second group.

Starting with the first group, we will issue the following commands:

```
. ssd init      variable names
. ssd set obs   values
. ssd set means values
. ssd set sd    values
. ssd set corr  values
```

We will first set the end-of-line delimiter to a semicolon because we are going to have some long lines. We will be entering SSD for 16 variables!

```
. #delimiter ;
delimiter now ;
. ssd init phyab1  phyab2  phyab3  phyab4
>          appear1 appear2 appear3 appear4
>          peerrel1 peerrel2 peerrel3 peerrel4
>          parrel1  parrel2  parrel3  parrel4 ;
```

Summary statistics data initialized. Next use, in any order,

```
ssd set observations (required)
It is best to do this first.
```

```
ssd set means (optional)
Default setting is 0.
```

```
ssd set variances or ssd set sd (optional)
Use this only if you have set or will set correlations and, even
then, this is optional but highly recommended. Default setting is 1.
```

```
ssd set covariances or ssd set correlations (required)
```

```
. ssd set obs 134 ;
(value set)
      Status:
              observations:  set
              means:        unset
              variances or sd:  unset
              covariances or correlations:  unset (required to be set)

. ssd set means
>      8.34 8.34 8.37 8.40 7.51 7.22 7.03 7.13
>      8.44 7.62 7.06 7.89 9.32 9.39 8.69 9.13 ;
(value set)
      Status:
              observations:  set
              means:        set
              variances or sd:  unset
              covariances or correlations:  unset (required to be set)

. ssd set sd
>      1.90 1.75 2.06 1.88 2.30 2.63 2.71 2.42
>      2.05 2.22 2.38 2.12 1.21 1.21 1.71 1.32 ;
(value set)
      Status:
              observations:  set
              means:        set
              variances or sd:  set
              covariances or correlations:  unset (required to be set)

. ssd set corr
>      1.0 \
>      .50 1.0 \
>      .59 .46 1.0 \
>      .58 .43 .66 1.0 \
>      .30 .27 .35 .46 1.0 \
>      .32 .34 .38 .39 .71 1.0 \
>      .38 .41 .43 .53 .68 .67 1.0 \
>      .23 .29 .33 .43 .61 .63 .73 1.0 \
>      .43 .32 .40 .42 .36 .34 .45 .42 1.0 \
>      .38 .40 .38 .49 .53 .61 .69 .59 .59 1.0 \
>      .27 .24 .41 .37 .43 .46 .57 .57 .61 .59 1.0 \
>      .43 .41 .37 .47 .51 .45 .63 .61 .59 .58 .65 1.0 \
>      .20 .14 .15 .18 .22 .21 .13 .03 .15 .19 .12 .14 1.0 \
>      .29 .18 .26 .20 .25 .29 .17 .25 .35 .23 .23 .28 .25 1.0 \
>      .37 .14 .34 .37 .34 .34 .35 .33 .42 .36 .39 .39 .53 .50 1.0 \
>      .13 .10 .16 .21 .33 .28 .23 .22 .23 .25 .23 .28 .46 .43 .59 1.0 ;
(value set)
      Status:
              observations:  set
              means:        set
              variances or sd:  set
              covariances or correlations:  set

. #delimit cr
delimiter now cr
```

We have now entered the data for the first group, and ssd reports that we have a fully set dataset.

Next we are going to add a second group by typing

```
. ssd addgroup grade
  (new group grade==2 added)

  The ssd set commands now modify the new group grade==2. If you need to
  modify data for grade==1, place a 1 right after the set. For example,

  . ssd set 1 means ...

  would modify the means for group grade==1.
```

The **ssd set** command now modifies the new group grade==2. If we needed to modify data for grade==1, we would place a 1 right after the **set**. For example,

```
. ssd set 1 means ...
```

We are not modifying data; however, we are now adding data for the second group. The procedure for entering the second group is the same as the procedure for entering the first group:

```
. ssd set obs   values
. ssd set means values
. ssd set sd    values
. ssd set corr  values
```

We do that below.

```
. #delimit ;
delimiter now ;
. ssd set obs 251 ;
  (value set for group grade==2)

  Status for group grade==2:
      observations:   set
      means:         unset
      variances or sd:  unset
      covariances or correlations:  unset (required to be set)

. ssd set corr
>   1.0 \
>   .31 1.0 \
>   .52 .45 1.0 \
>   .54 .46 .70 1.0 \
>   .15 .33 .22 .21 1.0 \
>   .14 .28 .21 .13 .72 1.0 \
>   .16 .32 .35 .31 .59 .56 1.0 \
>   .23 .29 .43 .36 .55 .51 .65 1.0 \
>   .24 .13 .24 .23 .25 .24 .24 .30 1.0 \
>   .19 .26 .22 .18 .34 .37 .36 .32 .38 1.0 \
>   .16 .24 .36 .30 .33 .29 .44 .51 .47 .50 1.0 \
>   .16 .21 .35 .24 .31 .33 .41 .39 .47 .47 .55 1.0 \
>   .08 .18 .09 .12 .19 .24 .08 .21 .21 .19 .19 .20 1.0 \
>   .01 -.01 .03 .02 .10 .13 .03 .05 .26 .17 .23 .26 .33 1.0 \
>   .06 .19 .22 .22 .23 .24 .20 .26 .16 .23 .38 .24 .42 .40 1.0 \
>   .04 .17 .10 .07 .26 .24 .12 .26 .16 .22 .32 .17 .42 .42 .65 1.0 ;
  (values set for group grade==2)

  Status for group grade==2:
      observations:   set
      means:         unset
      variances or sd:  unset
      covariances or correlations:   set
```

```
. ssd set sd      1.84 1.94 2.07 1.82 2.34 2.61 2.48 2.34
>                1.71 1.93 2.18 1.94 1.31 1.57 1.77 1.47 ;
(values set for group grade==2)

    Status for group grade==2:
                observations:    set
                        means:   unset
                variances or sd:    set
    covariances or correlations:    set

. ssd set means  8.20 8.23 8.17 8.56 7.41 7.00 7.17 7.40
>                8.81 7.94 7.52 8.29 9.35 9.13 8.67 9.00 ;
(values set for group grade==2)

    Status for group grade==2:
                observations:    set
                        means:    set
                variances or sd:    set
    covariances or correlations:    set

. #delimit cr
delimiter now cr
```

We could stop here and save the data in a Stata dataset. We might type

```
. save sem_2fmmby
```

However, we intend to use these data as an example in this manual and online. Here is what you would see if you typed `ssd describe`:

```
. ssd describe
Summary statistics data
Observations:          385
Variables:              16
```

Variable name	Variable label
phyab1	
phyab2	
phyab3	
phyab4	
appear1	
appear2	
appear3	
appear4	
peerrel1	
peerrel2	
peerrel3	
peerrel4	
parrel1	
parrel2	
parrel3	
parrel4	

```

Group variable: grade (2 groups)
Obs by group: 134, 251
```

We are going to label these data so that `ssd describe` can provide more information:

```
. label data "two-factor CFA"
. label var phyab1 "Physical ability 1"
. label var phyab2 "Physical ability 2"
. label var phyab3 "Physical ability 3"
. label var phyab4 "Physical ability 4"
. label var appear1 "Appearance 1"
. label var appear2 "Appearance 2"
. label var appear3 "Appearance 3"
. label var appear4 "Appearance 4"
. label var peerrel1 "Relationship w/ peers 1"
. label var peerrel2 "Relationship w/ peers 2"
. label var peerrel3 "Relationship w/ peers 3"
. label var peerrel4 "Relationship w/ peers 4"
. label var parrel1 "Relationship w/ parent 1"
. label var parrel2 "Relationship w/ parent 2"
. label var parrel3 "Relationship w/ parent 3"
. label var parrel4 "Relationship w/ parent 4"
. #delimit ;
delimiter now ;
. notes: Summary statistics data from
> Marsh, H. W. and Hocevar, D., 1985,
> "Application of confirmatory factor analysis to the study of
> self-concept: First- and higher order factor models and their
> invariance across groups", Psychological Bulletin, 97: 562-582. ;
. notes: Summary statistics based on
> 134 students in grade 4 and
> 251 students in grade 5
> from Sydney, Australia. ;
. notes: Group 1 is grade 4, group 2 is grade 5. ;
. notes: Data collected using the Self-Description Questionnaire
> and includes sixteen subscales designed to measure
> nonacademic traits: four intended to measure physical
> ability, four intended to measure physical appearance,
> four intended to measure relations with peers, and four
> intended to measure relations with parents. ;
. #delimit cr
delimiter now cr
```

We would now save the dataset.

To see `ssd describe`'s output with the data labeled, see [SEM] [Example 20](#).

## Reference

Marsh, H. W., and D. Hocevar. 1985. Application of confirmatory factor analysis to the study of self-concept: First- and higher order factor models and their invariance across groups. *Psychological Bulletin* 97: 562–582. <https://doi.org/10.1037/0033-2909.97.3.562>.

## Also see

[SEM] [Example 20](#) — Two-factor measurement model by group

[SEM] [ssd](#) — Making summary statistics data (sem only)

Stata, Stata Press, Mata, NetCourse, and NetCourseNow are registered trademarks of StataCorp LLC. Stata and Stata Press are registered trademarks with the World Intellectual Property Organization of the United Nations. StataNow is a trademark of StataCorp LLC. Other brand and product names are registered trademarks or trademarks of their respective companies. Copyright © 1985–2025 StataCorp LLC, College Station, TX, USA. All rights reserved.



For suggested citations, see the FAQ on [citing Stata documentation](#).