

Example 2 — Creating a dataset from published covariances

[Description](#)

[Remarks and examples](#)

[Reference](#)

[Also see](#)

Description

Williams, Eaves, and Cox (2002) publish covariances from their data. We will use those published covariances in [SEM] Example 3 to fit an SEM.

In this example, we show how to create the summary statistics dataset (SSD) that we will analyze in that example.

Remarks and examples

Remarks are presented under the following headings:

- Background
- Creating the SSD
- At this point, we could save the dataset and stop
- Labeling the SSD
- Listing the SSD

For more explanation, also see [SEM] Intro 11.

Background

In Williams, Eaves, and Cox (2002), the authors report a covariance matrix in a table that looks something like this:

	Affective		Miniscale		Cognitive		Miniscale
	1	2	...	5	1	2	...
Affective							
1	2038.035	1631.766	...	1721.830	659.795	779.519	...
2		1932.163	...	1688.292	702.969	790.448	...
.							
.							
5				2061.875	775.118	871.211	...
Cognitive							
1					630.518	500.128	...
.							
.							

Creating the SSD

```
. clear all
. ssd init a1 a2 a3 a4 a5 c1 c2 c3 c4 c5
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 216
(value set)
Status:
              observations:  set
              means:       unset
              variances or sd:  unset
              covariances or correlations:  unset (required to be set)
. #delimit ;
delimiter now ;
. ssd set cov 2038.035 \
>      1631.766 1932.163 \
>      1393.567 1336.871 1313.809 \
>      1657.299 1647.164 1273.261 2034.216 \
>      1721.830 1688.292 1498.401 1677.767 2061.875 \
>      659.795 702.969 585.019 656.527 775.118 630.518 \
>      779.519 790.448 653.734 764.755 871.211 500.128 741.767 \
>
>      899.912 879.179 750.037 897.441 1008.287 648.935 679.970
>      1087.409 \
>
>      775.235 739.157 659.867 751.860 895.616 550.476 603.950
>      677.865 855.272 \
>
>      821.170 785.419 669.951 802.825 877.681 491.042 574.775
>      686.391 622.830 728.674 ;
(values set)
Status:
              observations:  set
              means:       unset
              variances or sd:  set
              covariances or correlations:  set
. #delimit cr
delimiter now cr
```

Notes:

1. We used `#delimit` to temporarily set the end-of-line character to semicolon. That was not necessary, but it made it easier to enter the data in a way that would be subsequently more readable. You can use `#delimit` only in do-files; see [\[P\] #delimit](#).
2. We recommend entering SSD by using do-files. That way, you can edit the file and get it right.
3. We did not have to reset the delimiter. We could have entered the numbers on one (long) line. That works well when there are only a few summary statistics.

At this point, we could save the dataset and stop

We could save the dataset and stop right now if we wished:

```
. save sem_2fmm
file sem_2fmm.dta saved
```

Obviously, we can save the dataset anytime we wish. We know we could stop because `ssd status` tells us whether there is anything more that we need to define:

```
. ssd status
Status:
              observations:  set
              means:       unset
              variances or sd:  set
              covariances or correlations:  set
```

Notes:

1. The means have not been set. The authors did not provide the means.
2. `ssd status` would mention if anything that was not set was required to be set.

Labeling the SSD

If we were to use `ssd describe` to describe these data, the output would look like this:

```
. ssd describe
Summary statistics data
Observations:          216
Variables:              10
```

Variable name	Variable label
a1	
a2	
a3	
a4	
a5	
c1	
c2	
c3	
c4	
c5	

We can add labels and notes to our dataset:

```
. label data "Affective and cognitive arousal"
. label var a1 "affective arousal 1"
. label var a2 "affective arousal 2"
. label var a3 "affective arousal 3"
. label var a4 "affective arousal 4"
. label var a5 "affective arousal 5"
. label var c1 "cognitive arousal 1"
. label var c2 "cognitive arousal 2"
. label var c3 "cognitive arousal 3"
. label var c4 "cognitive arousal 4"
```

```
. label var c5 "cognitive arousal 5"
. #delimit ;
delimiter now ;
. notes: Summary statistics data containing published covariances
> from Thomas O. Williams, Ronald C. Eaves, and Cynthia Cox,
> 2 Apr 2002, "Confirmatory factor analysis of an instrument
> designed to measure affective and cognitive arousal",
> _Educational and Psychological Measurement_,
> vol. 62 no. 2, 264-283. ;

. notes: a1-a5 report scores from 5 miniscales designed to measure
> affective arousal. ;

. notes: c1-c5 report scores from 5 miniscales designed to measure
> cognitive arousal. ;

. notes: The series of tests, known as the VST II
> (Visual Similes Test II) were administered to 216 children
> ages 10 to 12. The miniscales are sums of scores of
> 5 to 6 items in VST II. ;

. #delimit cr
delimiter now cr
. ssd describe

Summary statistics data
Observations:          216          Affective and cognitive arousal
Variables:             10
                        (_dta has notes)

Variable name          Variable label
-----
a1                     affective arousal 1
a2                     affective arousal 2
a3                     affective arousal 3
a4                     affective arousal 4
a5                     affective arousal 5
c1                     cognitive arousal 1
c2                     cognitive arousal 2
c3                     cognitive arousal 3
c4                     cognitive arousal 4
c5                     cognitive arousal 5

. save sem_2fmm, replace
file sem_2fmm.dta saved
```

Notes:

1. You can label the variables and the data, and you can add notes just as you would to any dataset.
2. You save and use SSD just as you save and use any dataset.

Listing the SSD

```
. ssd list
Observations = 216
Means undefined; assumed to be 0
Variances implicitly defined; they are the diagonal of the covariance
matrix.
Covariances:
      a1      a2      a3      a4      a5      c1      c2
2038.035
1631.766 1932.163
1393.567 1336.871 1313.809
1657.299 1647.164 1273.261 2034.216
1721.83 1688.292 1498.401 1677.767 2061.875
659.795 702.969 585.019 656.527 775.118 630.518
779.519 790.448 653.734 764.755 871.211 500.128 741.767
899.912 879.179 750.037 897.441 1008.287 648.935 679.97
775.235 739.157 659.867 751.86 895.616 550.476 603.95
821.17 785.419 669.951 802.825 877.681 491.042 574.775
      c3      c4      c5
1087.409
677.865 855.272
686.391 622.83 728.674
```

Reference

Williams, T. O., Jr., R. C. Eaves, and C. Cox. 2002. Confirmatory factor analysis of an instrument designed to measure affective and cognitive arousal. *Educational and Psychological Measurement* 62: 264–283. <https://doi.org/10.1177/0013164402062002005>.

Also see

[SEM] [Example 3](#) — Two-factor measurement model

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

Stata, Stata Press, and Mata are registered trademarks of StataCorp LLC. Stata and Stata Press are registered trademarks with the World Intellectual Property Organization of the United Nations. StataNow and NetCourseNow are trademarks 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](#).

