Confirmatory factor analysis of ordinal data

Contents

1. Assumption of normality

While all the observed variables are continuous, maximum likelihood estimation has an underlying assumption of multivariate normality. With non-normal continuous data, ML produces relatively accurate parameter estimates, but the bias in chi-square and standard errors increases with non-normality.

While WLS estimation produces accurate parameter estimates, there is a tendency to underestimate standard errors and overestimate goodness-of-fit measures. Larger or more complex models, or greater nonnormality, also sometimes causes a failure of WLS to converge.

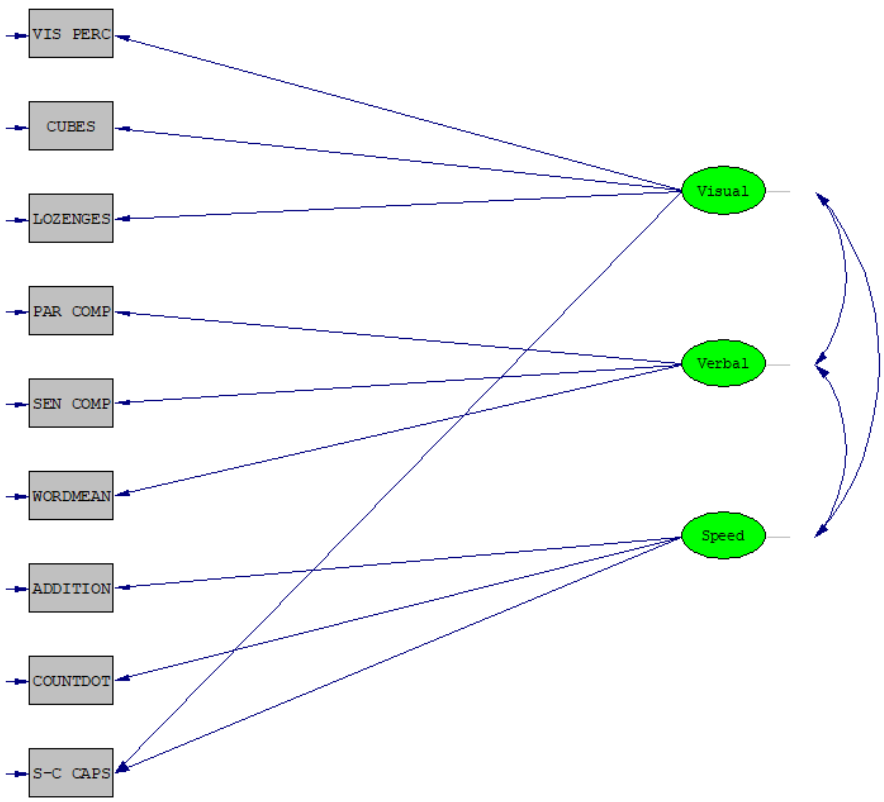
Modifying the full WLS approach to utilize only the diagonal elements of the asymptotic covariance matrix in an approach that became known as diagonally weighted least squares (DWLS). DWLS produces parameter estimates that appear to be less distorted by nonnormality than are ML estimates. Further modifications include using adjustment formulas to produce more accurate results, correcting standard errors and χ^2-values for bias due to nonnormality. This approach is commonly referred to as robust WLS or robust DWLS.

In this example (Holzinger & Swineford (1939)) nine variables were selected to measure three latent variables: Space, Verbal and Visual. The group of interest in this example consists of 145 eighth-grade children from the Grant-White school in Chicago.

To take a closer look at the impact of the non-normality of these variables on the estimates, standard error estimates and values of the goodness-of-fit statistics, we now fit a confirmatory factor analysis model to these data using the following three methods:

- ML estimation with normal theory standard errors
- WLS estimation, and
- ML estimation with robust standard errors

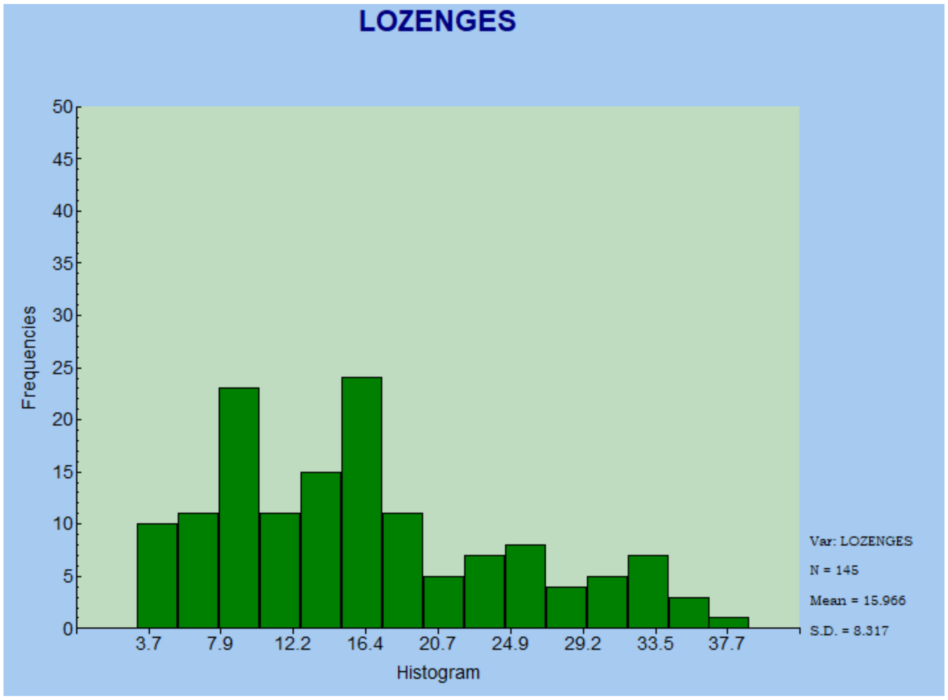
A conceptual path diagram of the model to be fitted is shown below.

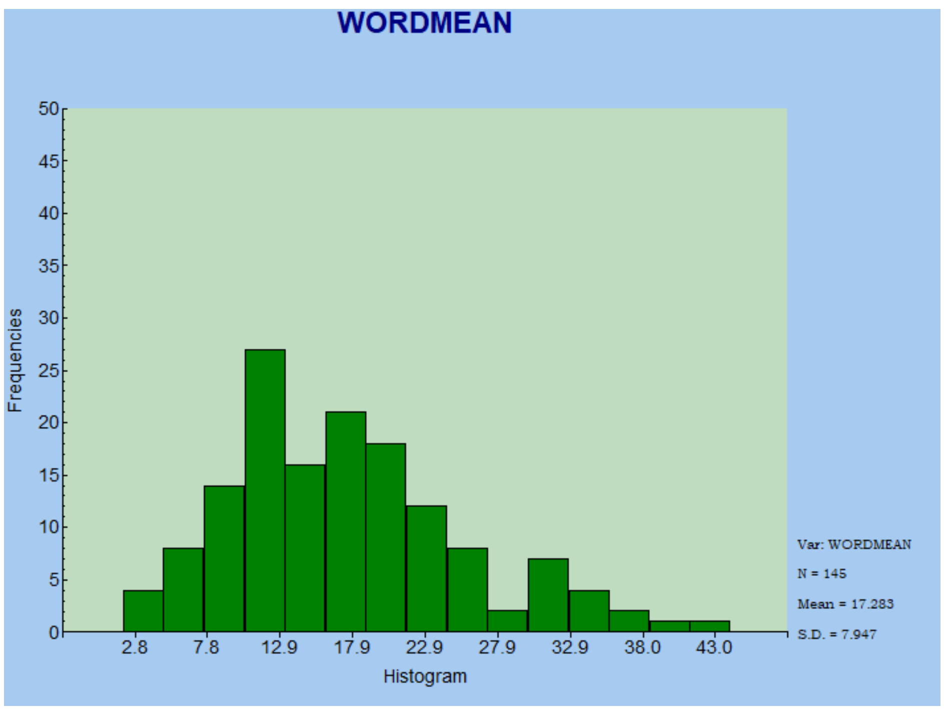


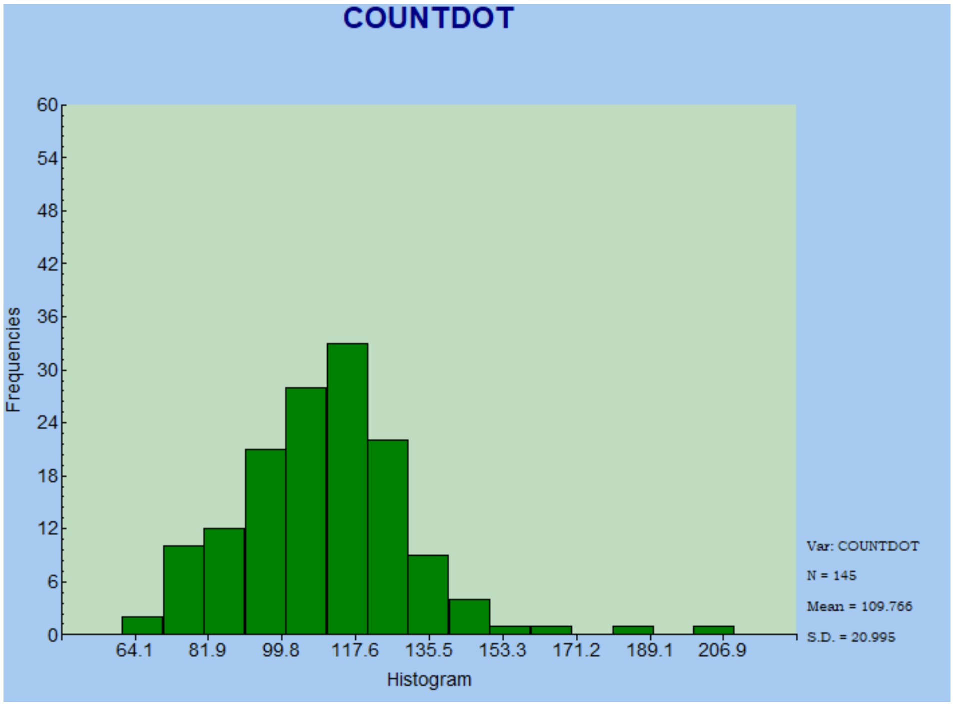
2. Data exploration

As a first step, we take a closer look at the distributions of the observed variables we intend to use. The data are contained in the LSF file **NPV.LSF**.

Bar charts for the variables LOZENGES, WORDMEAN and COUNTDOT are shown below. It is clear that these variables are not normally distributed.


LOZENGES
Frequencies
50
45
40
35
30
25
20
15
10
5
0
3.7
7.9
12.2
16.4
20.7
24.9
29.2
33.5
37.7
Histogram
Var: LOZENGES
N = 145
Mean = 15.966
S.D. = 8.317


WORDMEAN
Frequencies
50
45
40
35
30
25
20
15
10
5
0
2.8
7.8
12.9
17.9
22.9
27.9
32.9
38.0
43.0
Histogram
Var: WORDMEAN
N = 145
Mean = 17.283
S.D. = 7.947



A data screening confirms these suspicions:

Total Sample Size(N) = 145

Univariate Summary Statistics for Continuous Variables

Variable	Mean	St. Dev.	Skewness	Kurtosis	Minimum	Freq.	Maximum	Freq.
VISPERC	29.579	6.914	-0.119	-0.046	11.000	1	51.000	1
CUBES	24.800	4.445	0.239	0.872	9.000	1	37.000	2
LOZENGES	15.966	8.317	0.623	-0.454	3.000	2	36.000	1
PARCOMP	9.952	3.375	0.405	0.252	1.000	1	19.000	1
SENCOMP	18.848	4.649	-0.550	0.221	4.000	1	28.000	1
WORDMEAN	17.283	7.947	0.729	0.233	2.000	1	41.000	1
ADDITION	90.179	23.782	0.163	-0.356	30.000	1	149.000	1
COUNTDOT	109.766	20.995	0.698	2.283	61.000	1	200.000	1
SCCAPS	191.779	37.035	0.200	0.515	112.000	1	333.000	1

Test of Univariate Normality for Continuous Variables

	Skewness		Kurtosis		Skewness and Kurtosis	
Variable	Z-Score	P-Value	Z-Score	P-Value	Chi-Square	P-Value
VISPERC	-0.604	0.546	0.045	0.964	0.367	0.833
CUBES	1.202	0.229	1.843	0.065	4.842	0.089
LOZENGES	2.958	**0.003**	-1.320	0.187	10.491	**0.005**
PARCOMP	1.995	**0.046**	0.761	0.447	4.559	0.102
SENCOMP	-2.646	**0.008**	0.693	0.489	7.483	**0.024**
WORDMEAN	3.385	**0.001**	0.720	0.472	11.977	**0.003**
ADDITION	0.826	0.409	-0.937	0.349	1.560	0.458
COUNTDOT	3.263	**0.001**	3.325	0.001	21.699	**0.000**
SCCAPS	1.008	0.313	1.273	0.203	2.638	0.267

```
Relative Multivariate Kurtosis = 1.072

Test of Multivariate Normality for Continuous Variables

          Skewness                  Kurtosis           Skewness and Kurtosis

  Value  Z-Score P-Value     Value  Z-Score P-Value      Chi-Square P-Value
 ------  ------- -------   -------  ------- -------      ---------- -------
 11.733    5.426   0.000   106.098    3.023   0.003          38.579   0.000
```

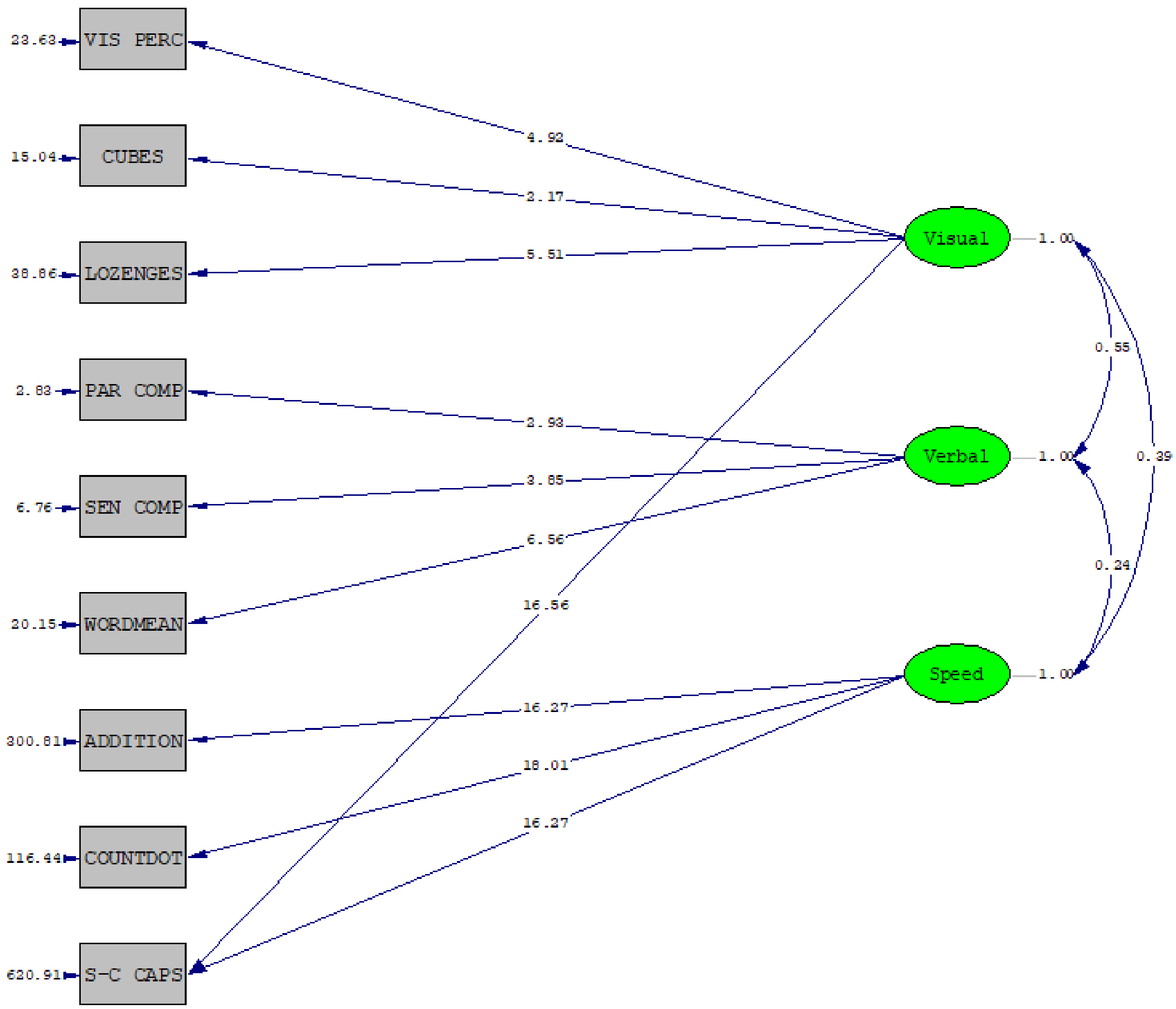
Based on these results, we anticipate that relatively accurate parameter estimates may be obtained using maximum likelihood estimation, but that increased bias in chi-square and standard errors may occur due to the non-normality in some of the observed variables.

3. Maximum likelihood estimation with normal theory standard errors

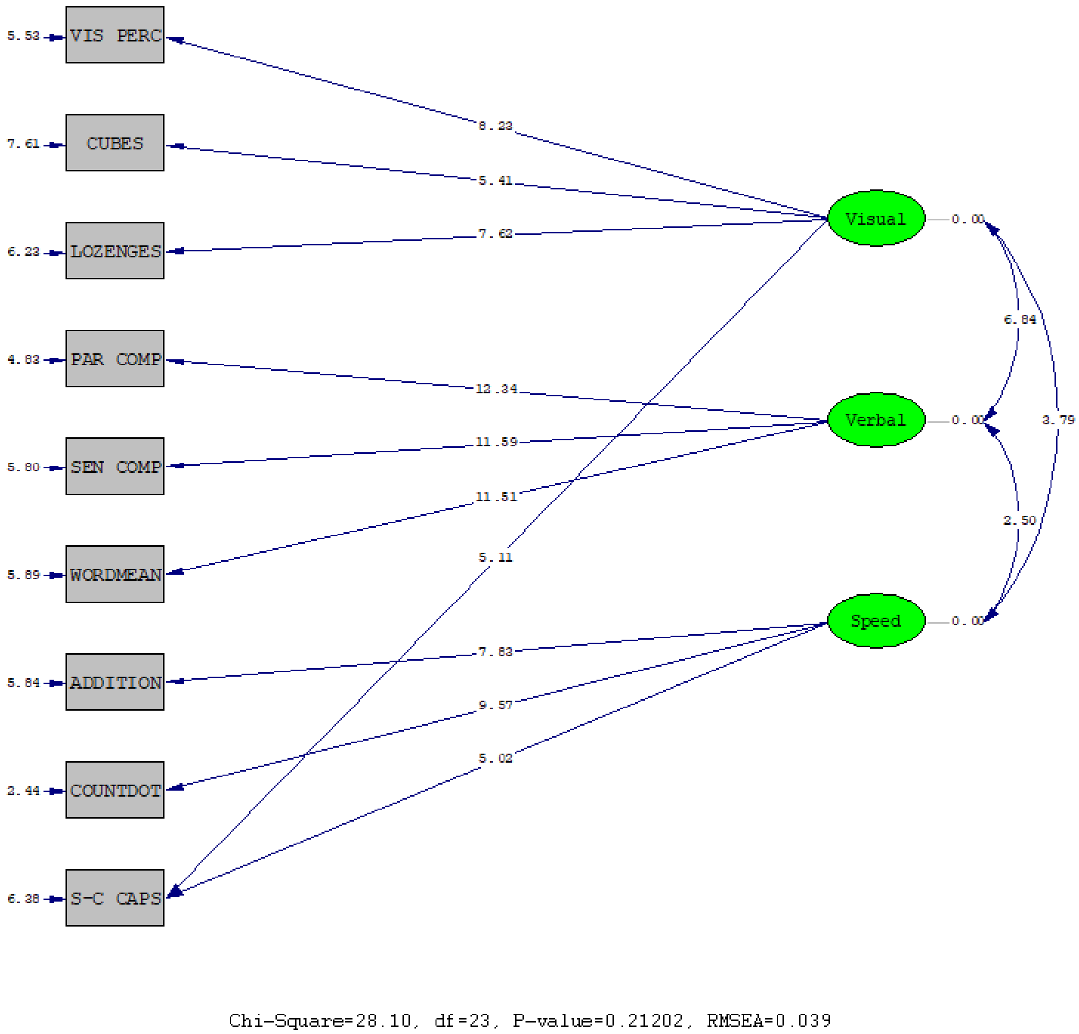
The SIMPLIS syntax for the first analysis is given in **hoswgw4.spl**, which can be found in the **SIMPLIS Examples\Factor Analysis and PCA** folder. The analysis is based on the covariance matrix saved in the file **HOSWGW.CM**. The SIMPLIS syntax below specifies the fitting of a confirmatory factor analysis with three latent variables.

```
Nine Psychological Variables - A Confirmatory Factor Analysis
Estimating Model by ML with Normal Theory Standard Errors
Observed Variables
   'VIS PERC' CUBES LOZENGES 'PAR COMP' 'SEN COMP' WORDMEAN
    ADDITION COUNTDOT 'S-C CAPS'
Covariance Matrix From File HOSWGW.CM
Sample Size 145
Latent Variables: Visual Verbal Speed
Relationships:
   'VIS PERC' - LOZENGES = Visual
   'PAR COMP' - WORDMEAN = Verbal
    ADDITION - 'S-C CAPS' = Speed
   'S-C CAPS' = Visual
Number of Decimals = 3
Print Residuals
Path Diagram
End of Problem
```

The estimates, fit statistics and *t*-values for this analysis are shown on the path diagrams below.



Chi-Square=28.10, df=23, P-value=0.21202, RMSEA=0.039



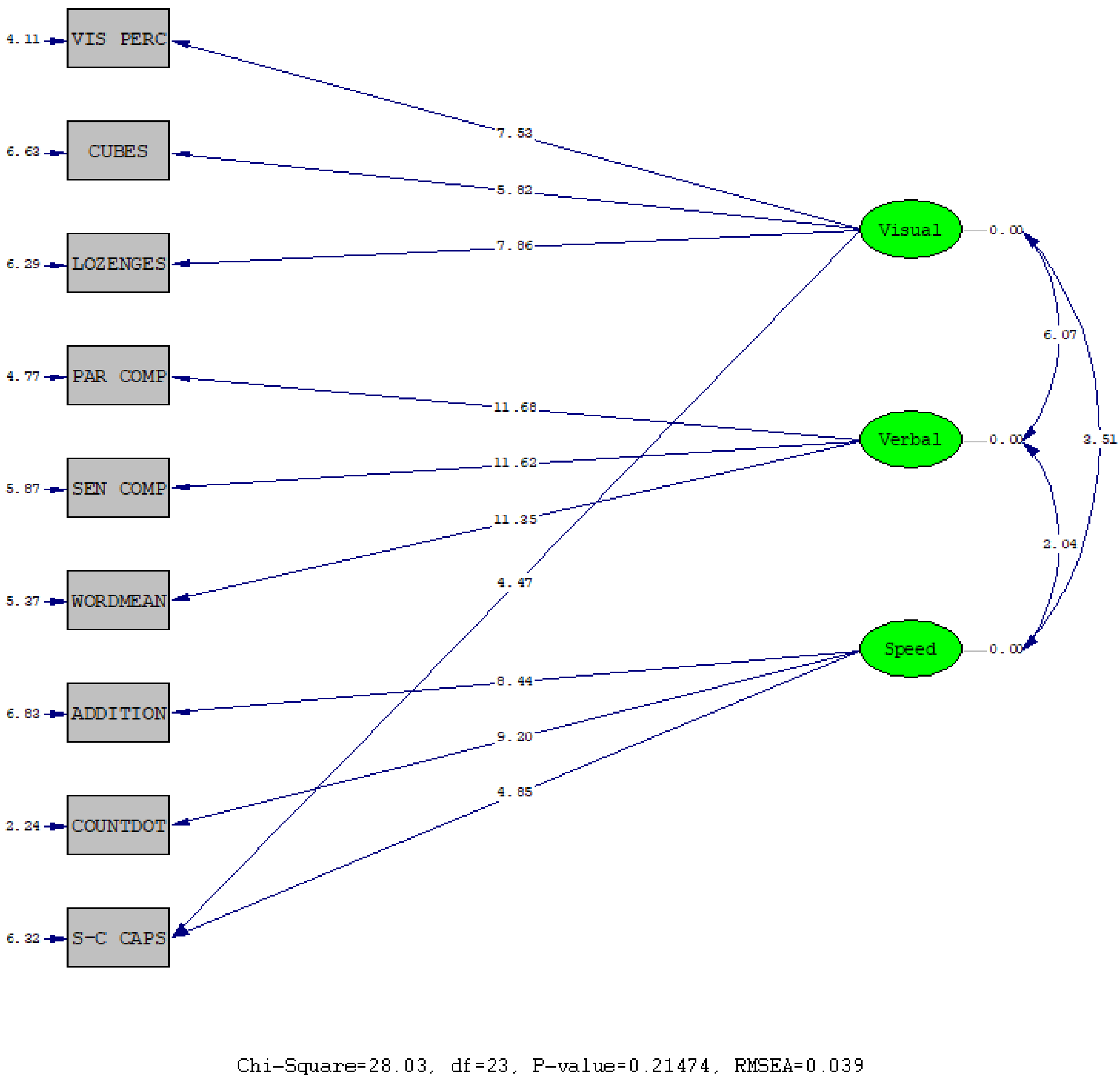
4. Maximum likelihood estimation with robust standard errors

The SIMPLIS syntax for the second analysis, using maximum likelihood estimation with robust standard errors, is shown below (**hoswgw5.spl**). The asymptotic covariance matrix is read in from the file **HOSWGW.ACC**. Maximum likelihood estimation is specified on the Method command line.

```
Nine Psychological Variables - A Confirmatory Factor Analysis
Estimating Model by ML with Robust Standard Errors
Observed Variables
   'VIS PERC' CUBES LOZENGES 'PAR COMP' 'SEN COMP' WORDMEAN
    ADDITION COUNTDOT 'S-C CAPS'
Covariance Matrix From File HOSWGW.CM
Asymptotic Covariance Matrix From File HOSWGW.ACC
Sample Size 145
Latent Variables: Visual Verbal Speed
Relationships:
   'VIS PERC' - LOZENGES = Visual
   'PAR COMP' - WORDMEAN = Verbal
   ADDITION - 'S-C CAPS' = Speed
```

```
   'S-C CAPS' = Visual
Number of Decimals = 3
Print Residuals
Method: Maximum Likelihood
Path Diagram
End of Problem
```

Inspection of the path diagram shows that the estimates are identical to those obtained for the maximum likelihood analysis considered in the previous section. We note differences in *t*-values, as illustrated on the path diagram below. Fit measures are essentially the same as obtained for the first analysis.

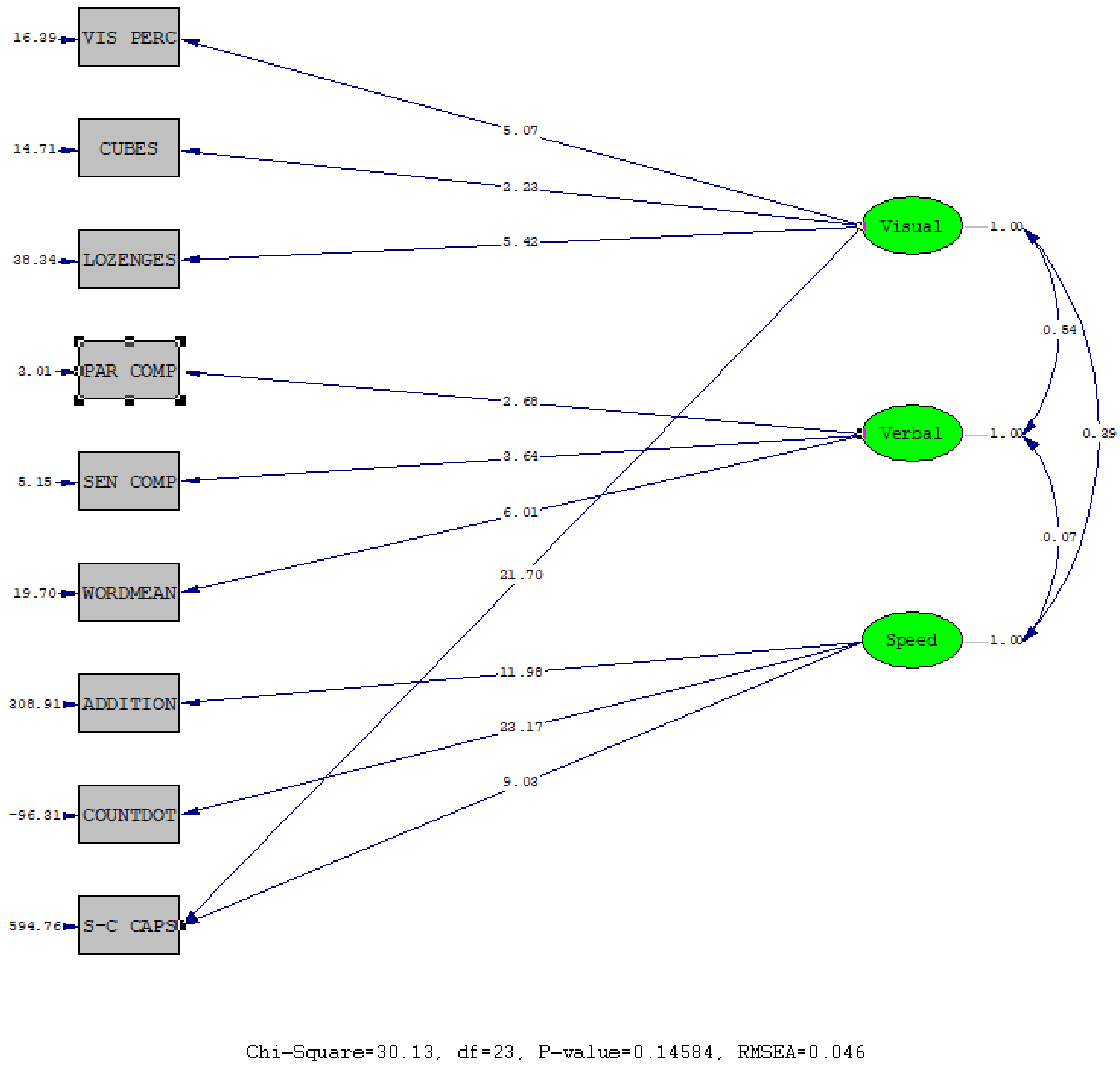


5. Weighted Least Squares estimation

The final analysis utilizes weighted least squares estimation. The only change in this syntax file (**hoswgw6.spl**), compared to the previous example, is in the Method command line where weighted least squares is specified.

```
Nine Psychological Variables - A Confirmatory Factor Analysis
Estimating Model by WLS
Observed Variables
   'VIS PERC' CUBES LOZENGES 'PAR COMP' 'SEN COMP' WORDMEAN
    ADDITION COUNTDOT 'S-C CAPS'
Covariance Matrix From File HOSWGW.CM
Asymptotic Covariance Matrix From File HOSWGW.ACC
Sample Size 145
Latent Variables: Visual Verbal Speed
Relationships:
   'VIS PERC' - LOZENGES = Visual
   'PAR COMP' - WORDMEAN = Verbal
    ADDITION - 'S-C CAPS' = Speed
   'S-C CAPS' = Visual
Number of Decimals = 3
Print Residuals
Method of Estimation: Weighted Least Squares
Path Diagram
End of Problem
```

Results for this analysis differ from the previous two analyses, both in estimates and *t*-values.



6. Comparing results

The table below summarizes the estimates, standard error estimates, *t*-values and values of selected fit statistics obtained for the three methods in the previous sections.

	Normal theory standard errors (method 1)			ML with robust standard errors (method 2)			Weighted Least Squares (method 3)		
Observed variable	**Visual**	**Verbal**	**Speed**	**Visual**	**Verbal**	**Speed**	**Visual**	**Verbal**	**Speed**
VIS PERC	4.917 (0.597) 8.234			4.917 (0.653) 7.534			5.073 (0.524) 9.703		

CUBES	2.173 (0.402) 5.408			2.173 (0.373) 5.818			2.229 (0.323) 6.909		
LOZENGES	5.506 (0.723) 7.616			5.506 (0.700) 7.862			5.420 (0.586) 9.249		
PAR COMP		2.926 (0.237) 12.341			2.926 (0.251) 11.676			2.676 (0.226) 11.850	
SEN COMP		3.855 (0.333) 11.590			3.855 (0.332) 11.620			3.637 (0.300) 12.120	
WORDMEAN		6.558 (0.570) 11.513			6.558 (0.578) 11.351			6.007 (0.535) 11.233	
ADDITION			16.272 (2.079) 7.826			16.272 (1.928) 8.441			11.981 (1.609) 7.448
COUNTDOT			18.010 (1.883) 9.565			18.010 (1.957) 9.202			23.168* (2.226) 10.410
S-C CAPS		16.559 (3.242) 5.108	16.274 (3.245) 5.015		16.559 (3.700) 4.475	16.274 (3.359) 4.845		9.025 (2.392) 7.364	21.704 (2.947) 3.773
Chi-square		28.10			28.03			30.13	
Degrees of freedom		23			23			23	
P-value		0.21202			0.21474			0.14584	
RMSEA		0.039			0.039			0.046	

*Negative error variance

Unfortunately, in the case of method 3, the weighted least squares estimation algorithm failed to obtain an admissible solution as indicated by the negative error variance estimate for COUNTDOT. As a result, it is not sensible to compare the corresponding estimates, standard error estimates, and the values of the goodness-of-fit statistics with those obtained with the other two methods.

As expected, we note identical parameter estimates for the first two methods. Comparing the standard errors estimated by the first two methods, we see that the standard error estimates of method 2 are in some cases larger and in some cases smaller than the standard error estimates obtained with the maximum likelihood method with normal theory standard errors. When we take a closer look at the standard error estimates for LOZENGES, WORDMEAN and COUNTDOT, we see an increase in the estimated standard errors of the latter two variables. The robust maximum likelihood Chi-square value is slightly smaller than the corresponding maximum likelihood value whilst the RMSEA point estimates are the same up to three decimal places.