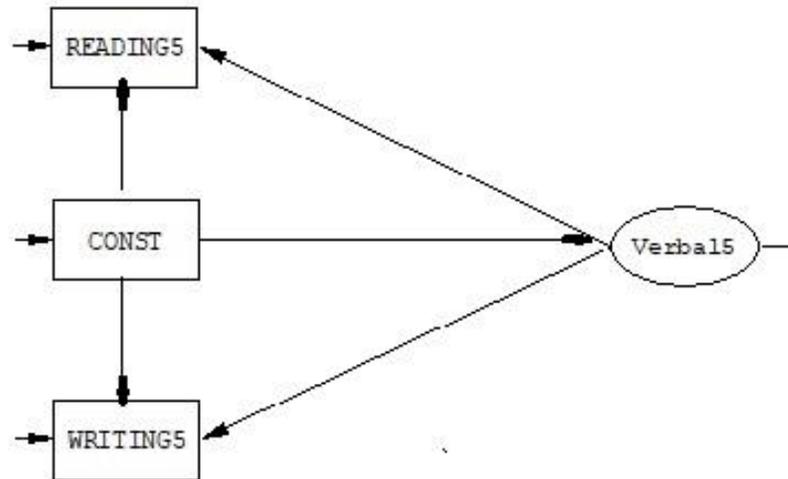# Regression models with latent variables: Regression of Verbal7 on Verbal5

Using the variables and data in the table below, take READING5 and WRITING5 to be indicators of a latent variable Verbal5 (Verbal ability at Grade 5), READING7 and WRITING7 to be indicators of a latent variable Verbal7 (Verbal ability at Grade 7) and estimate the regression of Verbal7 on Verbal5 for both groups.

Table 1: Means and covariance matrices for STEP Reading and writing

Boys Academic (N = 373)			
STEP Reading, Grade 5 — 281.349			
STEP Writing, Grade 5 — 184.219	182.821		
STEP Reading, Grade 7 — 216.739	171.699	283.289	
STEP Writing, Grade 7 — 198.376	153.201	208.837	246.069
Means — 262.236	258.788	275.630	269.075
Boys Non-academic (N = 249)			
STEP Reading, Grade 5 — 174.485			
STEP Writing, Grade 5 — 134.468	161.869		
STEP Reading, Grade 7 — 129.840	118.836	228.449	
STEP Writing, Grade 7 — 102.194	97.767	136.058	180.460
Means — 248.675	246.896	258.546	253.349

The measurement model is

The regression of Verbal7 on Verbal5 is of the form:

$$Verbal7 = \alpha + \gamma Verbal5 + z$$

Since the means of Verbal7 and Verbal5 are only determined relative to a fixed origin for both groups, it follows that we can set $\alpha = 0$ in the first group and estimate α in the second group. If the slope γ is the same in both groups, we can interpret α as an effect due to group differences. The syntax and data files can be found in the **SIMPLIS Examples** folder.

The input file for this analysis is **EX15A.SPL**:

```
Group BA: STEP Reading and Writing, Grades 5 and 7
Observed Variables: READING5 WRITING5 READING7 WRITING7
Covariance Matrix from File EX12.DAT
Means from File EX12.DAT
Sample Size: 373
Latent Variables: Verbal5 Verbal7
Relationships:
   READING5 = CONST + 1*Verbal5
   WRITING5 = CONST + (1)*Verbal5
   READING7 = CONST + 1*Verbal7
   WRITING7 = CONST + (1)*Verbal7
   Verbal7 = Verbal5

Group BNA: STEP Reading and Writing, Grades 5 and 7
Covariance Matrix from File EX12.DAT
Means from File EX12.DAT
Sample Size: 249
Relationships:
```

```
      Verbal5 = CONST
      Verbal7 = CONST + Verbal5
   Set the Error Variances of READING5 - WRITING7 free
   Set the Variance of Verbal5 free
   Set the Error Variance of Verbal7 free
   Path Diagram
   End of Problem
```

Here the line

```
Verbal7 = Verbal5
```

in the first group corresponds to the equation above with $\alpha = 0$ for the first group. The line

```
Verbal7 = CONST + Verbal5
```

in the second group corresponds to the equation with a different γ estimated in the second group. The line

```
      Verbal5 = CONST
```

in the second group will make the program estimate the mean of Verbal5 in the second group as in the previous examples. This mean is zero in group 1.

The two lines

```
Set the Variance of Verbal5 free
Set the Error Variance of Verbal7 free
```

in the second group indicate that we are not assuming the variance of Verbal5 to be the same in the two groups, neither are we assuming that the regression residual variance is the same.

The chi-square for this model is 10.11 with 6 degrees of freedom. This is similar to that in the example (the 13th in the set) where the same data were used. In fact, these two models are just two different parameterizations of the means, variances, and covariances in the table above.

```
  Verbal7 = 0.964*Verbal5, Errorvar.= 29.148, R² = 0.875
 Standerr  (0.0398)                      (6.837)
 Z-values   24.240                        4.263
 P-values   0.000                         0.000



  Verbal7 =  - 6.111  + 0.812*Verbal5, Errorvar.= 50.624, R² = 0.671
 Standerr    (1.101)  (0.0576)                     (9.848)
 Z-values    -5.551   14.089                        5.141
 P-values     0.000    0.000                        0.000
```

The two slopes are estimated as $\hat{\gamma}^{(1)} = 0.96$ (0.04) and $\hat{\gamma}^{(2)} = 0.81$ (0.06). Approximate 95 % confidence

intervals for the slopes would be $0.88 \le \hat{\gamma}^{(1)} \le 1.04$ and $0.69 \le \hat{\gamma}^{(21)} \le 0.93$. Since these overlap to some

extent, it is of interest to test the hypothesis $\hat{\gamma}^{(1)} = \hat{\gamma}^{(2)}$.

To do so, replace the line

```
Verbal7 = CONST + Verbal5
```

in the second group with

```
Verbal7 = CONST
```

as shown in example **EX15B.SPL**:

```
Group BA: STEP Reading and Writing, Grades 5 and 7
Observed Variables: READING5 WRITING5 READING7 WRITING7
Covariance Matrix from File EX12.DAT
Means from File EX12.DAT
Sample Size: 373
Latent Variables: Verbal5 Verbal7
Relationships:
   READING5 = CONST + 1*Verbal5
   WRITING5 = CONST + (1)*Verbal5
   READING7 = CONST + 1*Verbal7
   WRITING7 = CONST + (1)*Verbal7
   Verbal7 = Verbal5

Group BNA: STEP Reading and Writing, Grades 5 and 7
Covariance Matrix from File EX12.DAT
Means from File EX12.DAT
Sample Size: 249
Relationships:
   Verbal5 = CONST
   Verbal7 = CONST
Set the Error Variances of READING5 - WRITING7 free
Set the Variance of Verbal5 free
Set the Error Variance of Verbal7 free
Path Diagram
End of Problem
```

This means that α will be estimated as the coefficient of CONST as before. But since there is a path from Verbal5 to Verbal7 in group 1, this path will remain invariant over groups. In this way, the slope γ is

estimated to be the same in both groups. Had we included Verbal5 as a second variable on the right side, we would have estimated a different γ in the second group, as we did before.

The chi-square for this model is:

```
                  Global Goodness-of-Fit Statistics

 Degrees of Freedom for (C1)-(C2)                    7
 Maximum Likelihood Ratio Chi-Square (C1)            15.383 (P = 0.0314)
```

while for the previous model it is

```
Degrees of Freedom for (C1)-(C2)                     6
 Maximum Likelihood Ratio Chi-Square (C1)            10.130 (P = 0.1193)
```

The difference in chi-square is approximately 5.25 with one degree of freedom. Thus, the hypothesis of equal slopes cannot be rejected at the 1% level. The estimate of α is -4.60 with a t-value of -5.12.