Multiple groups: testing equality of regressions

In this example, we consider the problem of testing whether a regression equation is the same in several populations. Suppose that a dependent variable y and a number of explanatory variables $x_1, x_2, ..., x_q$ are observed in two or more groups. We are interested in determining the extent to which the regression equation

$$y = \alpha + \gamma_1 x_1 + \gamma_2 x_2 + \cdots + + \gamma_q x_q + z$$

Is the same in different groups. We say that the regressions are:

- Equal if $\alpha, \gamma_1, \gamma_2, ..., \gamma_q,$ are the same in all groups
- Parallel if $\gamma_1, \gamma_2, ..., \gamma_q,$ are the same in all groups.

Normally the covariance matrix of the *x*-variables is not expected to be the same across groups and often one finds that the intercept terms differ between groups. It may also be the case that *only some* of the regression coefficients are the same across groups.

Sörbom (1976) gave the covariance matrices shown in Table 1. These are based on scores on the ETS Sequential Test of Educational Progress (STEP) for two groups of boys who took the test in both Grade 5 and Grade 7. The two groups were defined according to whether or not they were in the academic curriculum in Grade 12. We will use this data to demonstrate how one can test the equality of various regressions.

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

We begin by typing the data and saving it for both groups in file **EX12.DAT** in the form of covariance matrices and means. It should look like this:

```
281.349
184.219 182.821
216.739 171.699 283.289
198.376 153.201 208.837 246.069
262.236 258.788 275.630 269.075
174.485
134.468 161.869
129.840 118.836 228.449
102.194  97.767 136.058 180.460
248.675 246.896 258.546 253.349
```

As a first example, we consider the testing of equal regressions of STEP Reading at Grade 7 on STEP Reading and Writing at Grade 5.

Writing at Grade 7 is not involved in this example. For example, we may want to predict STEP Reading at Grade 7 from STEP Reading and Writing at Grade 5 and to see if the prediction equation is the same for both groups. A path diagram is shown below, where CONST is a variable which is constant equal to 1 for every case. The intercept α in the regression equation is the coefficient of CONST.
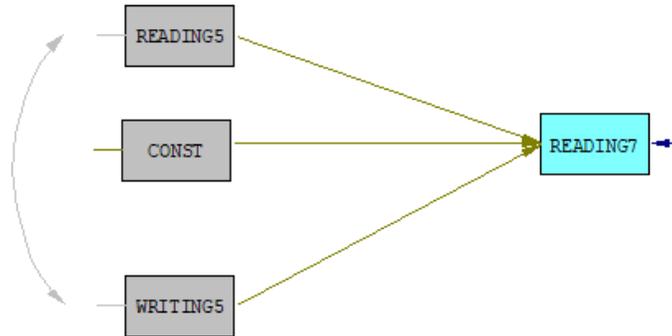
The input file for this example is **EX12A.SPL** (see the **SIMPLIS Examples** folder):

```
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
Equation: READING7 = CONST READING5 WRITING5

Group BNA: STEP Reading and Writing, Grades 5 and 7
Covariance Matrix from File EX12.DAT
Means from File EX12.DAT
Sample Size: 249
Set the Error Variance of READING7 Free
Path Diagram
End of Problem
```

Note the following:

- Four variables are read but only three variables are used in the model. The program automatically selects the variables for analysis on the basis of the variables included on the Equation line.
- The intercept α in the regression equation is the regression coefficient of the variable CONST. The line

  ```
  Equation: READING7 = CONST READING5 WRITING5
  ```

  is interpreted in the usual way, i.e., three coefficients α, γ_1, and γ_2 will be estimated. In order to estimate the intercept α, the means of the observed variables must be provided. If only covariance matrices are given, α will be zero.

- The line

  ```
  Set the Error Variance of READING7 Free
  ```

is needed in the second group; otherwise the regression error variance will be assumed to be the same in both groups.

The test of equal regressions gives a chi-square of

```
Degrees of Freedom for (C1)-(C2)                        3
 Maximum Likelihood Ratio Chi-Square (C1)                35.787 (P = 0.0000)
```

To test for parallel regressions rather than equal regressions, one must allow α to be different in the two groups. This is done by adding the line

```
Set the Path from CONST to READING7 Free
```

in the second group (see **EX12B.SPL**). Another way of specifying the same thing is to include the relationship

```
READING7 = CONST
```

in the second group. The meaning of this is that α, the coefficient of CONST, will be reestimated in the second group, while the remaining part of the regression equation remains the same as in group 1, i.e., the coefficients of READING5 and WRITING5 will be the same in both groups.

The test of parallel regressions gives a chi-square of 3.31 with two degrees of freedom. Thus, it is evident that the regressions are parallel but not equal.

As a second example, consider estimating two regression equations simultaneously in two groups and testing whether both regressions are parallel in the two groups. With the same data as in the previous example, we may estimate the regressions of READING7 and WRITING7 on READING5 and WRITING5 (**EX12C.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
Equations: READING7 - WRITING7 = CONST READING5 WRITING5
Set the Error Covariance between READING7 and WRITING7 free

Group BNA: STEP Reading and Writing, Grades 5 and 7
Covariance Matrix from File EX12.DAT
Means from File EX12.DAT
Sample Size: 249
Equations: READING7 - WRITING7 = CONST
Set the Error Variances of READING7 - WRITING7 Free
Set the Error Covariance Matrix of READING7 and WRITING7 Free
```

```
Path Diagram
End of Problem
```

- The line

  ```
  Set the Error Covariance between READING7 and WRITING7 free
  ```

  in the first group, specifies that the two error terms z_1 and z_2 may be correlated, i.e., we don't believe that READING5 and WRITING5 will account for the whole correlation between READING7 and WRITING7.

- The line

  ```
  Equations: READING7 - WRITING7 = CONST
  ```

  in the second group, specifies that the intercept terms for the second group do not have to equal those of the first group.

- The line

  ```
  Set the Error Variances of READING7 - WRITING7 Free
  ```

  in the second group specifies that the variances of z_1 and z_2 in the second group are not constrained to be equal to those of the first group.

- The line

  ```
  Set the Error Covariance Matrix of READING7 and WRITING7 Free
  ```

  in the second group specifies that the covariance between z_1 and z_2 in the second group is not constrained to be equal to that of the first group.

The chi-square for this model is

```
Degrees of Freedom for (C1)-(C2)                    4
 Maximum Likelihood Ratio Chi-Square (C1)            9.025 (P = 0.0605)
```

Another way of modeling these data is to take measurement error in the observed variables into account and to regard Reading and Writing as two indicators of a latent variable Verbal Ability, say, and to estimate the regression of Verbal Ability at Grade 7 on Verbal Ability at Grade 5. This will be considered elsewhere, after exploring how to estimate the means of latent variables.