Testing Research Hypotheses with Structural Equation Models: An Illustration with PROC CALIS

Mike W.-L. Cheung
Department of Psychology
The University of Hong Kong

- Paper presented at the SAS Academic User Group Conference 2005, Hong Kong

Introduction

- Testing statistical hypotheses is necessary for nearly all serious research

- Apart from obtaining results directly from statistical packages, researchers may need to conduct “non-standard” analyses sometimes.

Two examples

- Ex. 1: Comparing dependent correlation coefficients
- Ex. 2: Comparing equality of the beta coefficients in a multiple regression

Structural equation modeling (SEM)

- SEM is a general statistical framework for many multivariate statistics (Bollen, 1989), such as
  - Regression analysis
  - Path analysis
  - Factor analysis
  - ANOVA to MANCOVA
  - Canonical correlation analysis
  - Growth curve modeling
  - Meta-analytic structural equation model

- Many of the techniques we used in social sciences can be formulated in SEM
- SEM provides statistical tests for linear and non-linear constraints
- This means that we can test our complicated research hypotheses with SEM easily.

Example data

Table 1: Example data on three variables (N=100)

	LS	JS	HS
Life satisfaction (LS)	1.00		
Job satisfaction (JS)	0.33	1.00	
Home-life satisfaction (HS)	0.55	0.31	1.00
Standard deviations	1.97	1.87	1.50

Example 1: Testing dependent correlation coefficients

- For example,
  - Corr between LS and JS = .33
  - Corr between LS and HS = .55

- Is the difference statistically significant?

- *Note*. It is necessary to handle the dependence in testing these two correlations

Method proposed by Olkin and Finn (1990)

$$Z_1 = \frac{r_{ij} - r_{kl}}{\sqrt{\mathrm{var}(r_{ij}) + \mathrm{var}(r_{kl}) - 2\mathrm{cov}(r_{ij}, r_{kl})}}, \qquad (1)$$

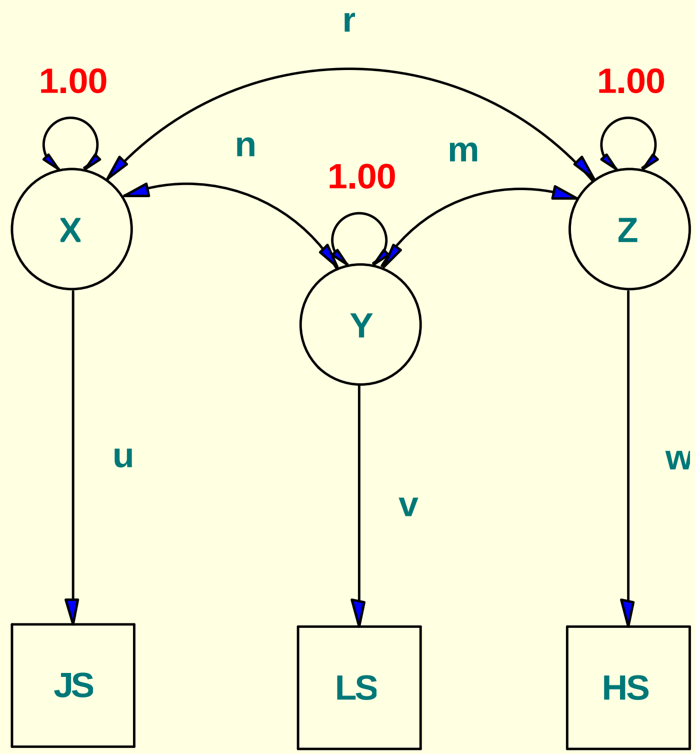
$$\mathrm{var}(r_{ij}) = \frac{(1 - r_{ij}^{\ 2})^2}{n} \text{ and}$$

$$\mathrm{cov}(r_{ij}, r_{kl}) = [0.5 r_{ij} r_{kl} (r_{ik}^{\ 2} + r_{il}^{\ 2} + r_{jk}^{\ 2} + r_{jl}^{\ 2}) + r_{ik} r_{jl} + r_{il} r_{jk} - (r_{ij} r_{ik} r_{il} + r_{ji} r_{jk} r_{jl} + r_{ki} r_{kj} r_{kl} + r_{li} r_{lj} r_{lk})] / n$$

where r_{ij} is the sample correlation coefficients of the ith and jth variables and n is the sample size.

- In our example, Z_1 is 2.17, p = .030. This suggests that the correlation between LS and HS is larger than the correlation between LS and JS.

SEM approach by Cheung and Chan (2004)



- We impose the equality on *m* and *n*.
- The result is nearly the same as the one with Equation 1: $\chi^2(1)=4.82$, $p=.028$.

SAS code: Testing dependent correlation coefficients

```
PROC CALIS Covariance DATA = Data_Table1;
        LINEQS
        LS = P_LS (1.97) F_LS,
        JS = P_JS (1.87) F_JS,
        HS = P_HS (1.50) F_HS;
* Set variances of variables;
        STD
        F_LS = 1.0,
        F_JS = 1.0,
        F_HS = 1.0;
* Set correlation between LS and JS equal correlation between LS and HS: C_Equal;
        COV
        F_LS F_JS = C_Equal,
        F_LS F_HS = C_Equal,
        F_JS F_HS = C_JS_HS;
* Select variables from the data;
        VAR LS JS HS;
RUN;
```

Example 2: Testing equality of beta coefficients in multiple regression

- Multiple regression is one of most frequently used statistics in behavioral sciences

- For example, LS = .178 + .186JS + .650HS, R^2=.331, p < .001

- It is of interest to test whether the beta coefficients are the same or not

- *Note*. For a meaningful comparison, the scales of JS and HS should be comparable.

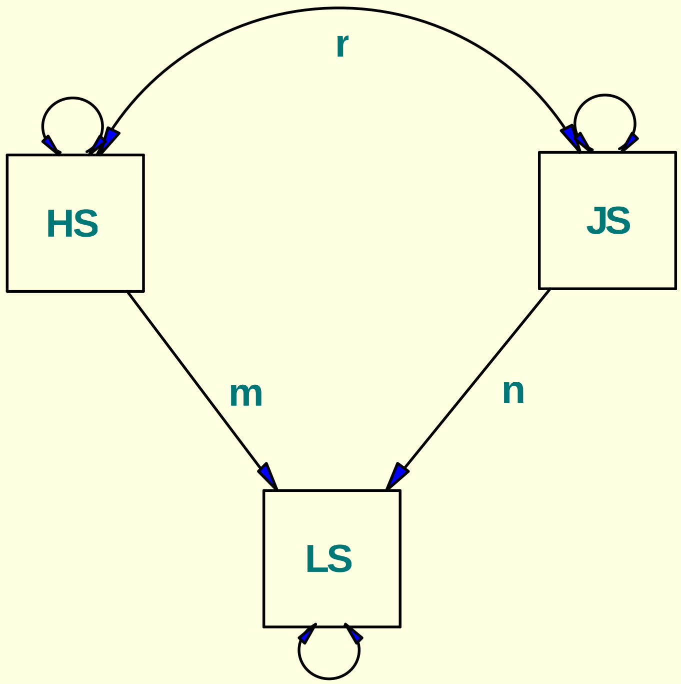
A large sample approximation

$$Z_2 = \frac{B_1 - B_2}{\sqrt{\text{var}(B_1) + \text{var}(B_2) - 2\text{cov}(B_1, B_2)}} \quad (2)$$

where B_1 and B_2 are the beta coefficients, $\text{var}(B_1)$ and $\text{var}(B_2)$ are the asymptotic variances of B_1 and B_2, and $\text{cov}(B_1, B_2)$ is the asymptotic covariance of B_1 and B_2.

- In our example, Z_2 is -2.76, p = .0057. This suggests that the regression coefficient for HS is larger than the regression coefficient for JS.

An SEM approach



- We impose the equality on *m* and *n*.
- The result is nearly the same as the one with Equation 2: $\chi^2(1)=7.52$, $p=.006$.

SAS code: Testing dependent correlation coefficients

```
PROC CALIS Covariance DATA = Data_Table1;
* Constrain the beta coefficients the same: P_EQ;
    LINEQS
    LS = P_EQ JS + P_EQ HS + E_LS;
* Set variances of variables;
    STD
    E_LS = VAR_LS,
    JS = VAR_JS,
    HS = VAR_HS;
* Covariances among variables;
    COV
    JS HS = C_JS_HS;
* Select variables from the data;
    VAR LS JS HS;
RUN;
```

Conclusion

- SEM is flexible for many multivariate statistics we encountered in typical research settings
- Advantages of PROC CALIS:
  - Integrated in SAS
- Disadvantages of PROC CALIS:
  - Single-group analysis currently

Selected references

- Bollen, K. A. (1989) *Structural equations with latent variables*. New York: Wiley.

- Cheung, M. W. L. and Chan, W. (2004) Testing dependent correlation coefficients via structural equation modeling. *Organizational Research Methods* 7, 206-223.

Thank you!