Latent variable modeling with applications to education assessment and NSF-REU projects for engineering students

Prof. Tak Cheung

Tak Cheung, Ph.D., professor of physics, teaches in CUNY Queensborough Community College. He also conducts research and mentors student research projects.

Dr. sunil Dehipawala, Queensborough Community College

Sunil Dehipawala received his B.S. degree from University of Peradeniya in Sri Lanka and Ph.D from City University of New York. Currently, he is working as a faculty member at Queensborough Community College of CUNY.

Dr. Rex Taibu

Dr. Rex Taibu has taught studio physics classes for several years. His teaching experience has shaped his research focus. Currently, Dr. Taibu is actively engaged in

1) promoting scientific inquiry attitudes in students through designing, implementing, and assessing innovative inquiry based physics labs.

2) conducting research regarding the role of language in conceptual understanding.

3) exploring cosmic rays (detection, data collection, and analysis).

Latent variable modeling with applications to education assessment and NSF-REU projects for engineering students

Vazgen Shekoyan, Raul Armendariz, Sunil Dehipawala, Rex Taibu, George Tremberger, David Lieberman, and Tak Cheung

CUNY Queensborough Community College Physics Department

Abstract

The latent variable modeling technique has been used to detect the presence of internal force in high impact projects. The technique can also be implemented in undergraduate research experiments where latent factors, present in simulated data, may be detected as part of a discovery experience. The example presented here uses the accelerations and weights of two masses in motion to illustrate the method. The learning objectives include the understanding of methods involving regression in contrast to those based on covariance, the numeric computation of pertinent coefficients, the exclusion of proposed causative models with relatively low loading values, and the understanding that the accepted causative model does not prove causality. In education assessment, latent variable modeling can be applied to pre-score and post-score data for tasks performed by different classes of students. The application of the educational research performed by Queensborough Community College's Physics faculty to NSF-REU projects for engineering students is discussed.

Keywords

Latent variable modeling, internal force, physics education assessment, REU projects

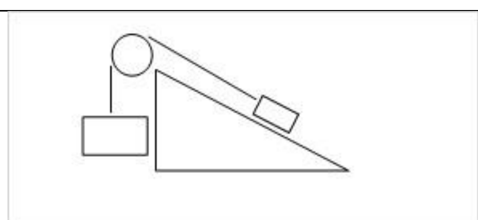
Introduction

Most of our community college STEM students are first generation college students. The students need advisement with concrete examples when professors describe different course requirements for the various majors. It has been a challenge to illustrate what kind of courses they would face after transferring to senior colleges. The latent variable method in discovering the presence of internal force in a physical situation would help STEM students to get a deeper understanding of the difference between a career in chemistry or one in chemical engineering. The latent variable approach has been used in chemical engineering and there is an online free textbook by Professor Kevin Dunn of McMaster University that describes it [1]. The latent variable approach in industrial electronics, brain computer interface applications, and structural equation modeling of power quality event data in electrical engineering have been published [2, 3, 4]. The application of the latent variable approach in internal force problems in physics would promote the implementation of discovery exercises using simulated data. The learning objectives include the understanding of methods involving regression in contrast to those based on covariance, the numeric computation of the pertinent coefficients, the exclusion of proposed causative models with relatively low loading values, and the understanding that the accepted causative model does not prove causality.

Regression analysis, which requires an independent variable, has been an important data analysis procedure in introductory physics lab for engineering students. The example presented here uses the accelerations and weights of two masses in motion to illustrate the methods involving regression in contrast to those based on covariance. On the other hand, education assessment data can also be studied using regression analysis. For example the Force Concept Inventory pre-score versus post-score data can be analyzed, and the regression slope has been interpreted as related to the learning gain after a semester of mechanics. The pre-score and post-score data can also be studied with latent variable modeling. The free student version of the modeling program, LISREL [5]. A quick start step by step instruction set for running LISREL on Windows is described in this paper, using a simulated dataset of force and acceleration. The extension of the latent variable modeling technique to our community college NSF-REU projects is also described.

The Internal Force Example

A standard internal force problem of two blocks connected by a string over a pulley is shown in Figure 1. The tensions on either side of the pulley are not equal because the angular acceleration of the pulley is affected by the torque. The initial parameter values are: 10-kg incline-mass, 50-kg hanging-mass, pulley moment of inertia 20-kg-m-sq, pulley radius 0.3-m, coefficient of kinetic friction 0.2, and 35-degree slope angle. The system acceleration value was calculated as 1.47 m/s/s. Changing the hanging-mass values or the incline-mass values would yield various acceleration values using Newton's law of motion. Adding random values would simulate lab data with noise, suitable for illustrating regression and covariance based methods. Note that without the noise addition, the data correlation would be perfect with coefficient of determination R-sq = 1. The random noise values were added using the RAND function in Microsoft Excel and the data was saved in CSV-format, suitable for running the LISREL software.



	A	B	C	D	E
1	Big-W	(grav)-a	small-W	a+(g-eff)	
2	499.5246	8.302593	101.5501	7.032437	
3	559.5721	8.288068	92.51064	7.145474	
4	566.4406	7.978542	85.70437	7.15479	
5	592.8642	7.931109	84.55598	7.133034	
6	649.4248	7.673328	80.8857	7.148694	
7	731.3791	7.59689	74.46468	7.194146	
8					

Figure 1: Left: A schematic of the studied inclined plane problem with two blocks and a pulley. Right: Using Newton's law of motion, system acceleration values were calculated using different weight values of the mass while keeping the other mass value constant. Noise was added to simulate lab data. Column-A and Column-B has R-sq = 0.84. Column-C and Column-D has R-sq = 0.80.

The discovery exercise was conducted by assigning the latent variable "Internal force 1" with observables "Big Weight (hanging weight)" and "gravity – acceleration (9.8 m/s/s – acceleration)". The latent variable "Internal force 2" was assigned with observables "Small Weight (incline-weight)" and "g-eff + acceleration (9.8* sin-angle m/s/s + acceleration)". The modeling result is shown in Figure 2. The relatively low loading value of 0.2 from "Internal force 2" to "Internal force 1" could suggest the presence of a third latent variable. A trial model with an additional latent variable due to torque was calculated using the data in Figure 1, but with interpolation so that there were 11 data rows. Two data columns were added to mimic the differential tension (Column-A – Column-C) and pulley acceleration ((Column-B + Column-D)*5), although Column-A-B data are simulated for constant incline-mass and Column-C-D data are simulated for constant hanging mass. The modeling result showed a relatively high loading value due to torque, as seen in Figure 3.

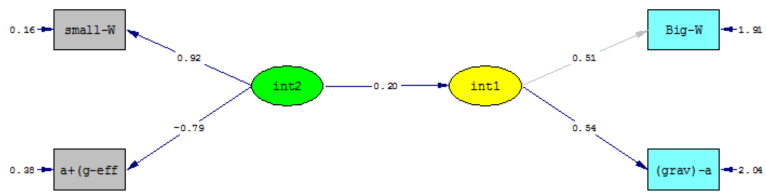


Figure 2: The result for the internal force model. The internal force on the hanging mass is "int1". The internal force on the incline-mass is "int2.

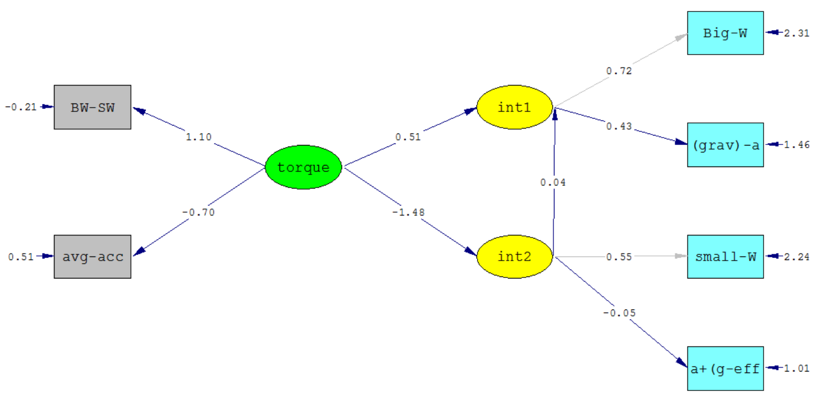


Figure 3: The result for the internal force model with 3 latent variables. The pulley effect was modeled as the third latent variable.

The step by step procedure for using the LISREL student version is as follows: (1) Calculate the data and put in the random noise as columns in a spreadsheet, with the first row containing variable names, then save in CSV format, and put the CSV file in a folder. (2) Open LISREL, select File/import data, open the CSV file and save as a LISREL system file. The computer would display the variable names. (3) Select File/new/path/, give a name for the path diagram of the postulated latent variables. (4) Select Setup/variable, select add-variable, select the system file. Browse for the system file, the computer will display the variable names, and select add-latent, "int1" and "int2" , in this case. (5) Drag the items from the left panel onto the diagram. (6) Select basic model, standard solution, just above the path diagram. (7) Select Setup/ build Lisrel syntax and then select File/run.

The Education Assessment

The assessment rubric of latent variable modeling of internal force as a discovery exercise is shown in Table 1. Students need a clear understanding of their expected deliverables.

Participant Deliverable	Highly Competent	Competent	Needs Improvement
Regression coefficient computation and interpretation (20%)	Provided clear and correct calculations of the regression coefficients in the given data sample.	Provided correct calculations of the regression coefficients but made one mistake	Contained two or more mistakes in the calculations.
Covariance value computation and interpretation (20%)	Provided clear and correct calculations of the covariance values in the given data sample.	Provided correct calculations of the covariance values but made one mistake.	Contained two or more mistakes in the calculations
Testing path diagram models (30%)	Provided correct results of three possible path diagrams.	Provided correct results of two possible path diagrams.	Only one of the three path diagram results was correct. Or all of the three path diagram results were wrong
Loading factor value interpretation (20%)	Provided a clear and correct interpretation of the loading factor values in each path diagram.	Provided a clear and correct interpretation of the loading factor values in any two path diagram.	Only one of the three path diagram loading value results was correctly interpreted. Or all path diagram loading factor results were misinterpreted.
Accepted causative model interpretation (10%)	Provided clear and correct numeric based reasons for rejecting the other path diagram models.	Provided one clear and correct numeric based reason for rejecting the other path diagram models.	Provided no numeric based reason for rejecting the other path diagram models.

Table 1: Latent variable modeling assessment rubric. The participants are students. Scoring could be performed when assigning Highly Competent = 1, Competent = 0.8 and Needs Improvement = 0.6.

Besides rubric-based score data, the pre-score and post-score data across time can also be used to assess learning skills. An example of education assessment in terms of latent variables is shown in Figure 4. The "intent" latent variable was assigned with the observables of pre-score and post-score, obtained from an assessment task in solving a physics problem. The "improve" latent variable was assigned with the observables of wri-1 and wri-2 data from an assessment task in lab report writing. The model could serve as an assessment of "learning with intent".

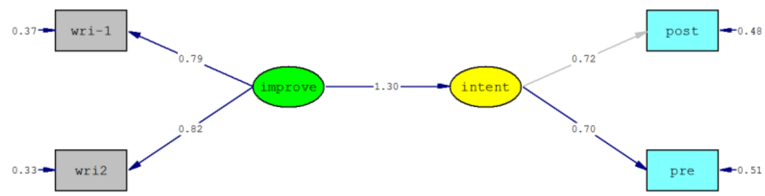


Figure 4: The result for the "learning with intent "latent variable model.

Discussion

The pre-score and post-score data collected for assessment can be studied with the latent variable technique so as to broaden the analysis scope beyond the regression method. The studying of learning gain as a latent variable would be able to further elucidate the various gain functions including post-score/pre-score relative to a baseline of zero, (post-score - pre-score)/(100 - pre-score) relative to full saturation, etc. The confirmatory factor analysis usually serves as a precursor, which is an effective stepping stone for community college students in the learning of latent variable modeling. There is at least one Microsoft Excel 2003 application for confirmatory factor analysis in the literature, which contains an example with one latent variable loading onto six data columns [6]. The confirmatory factor analysis exercise was found to be very suitable for community college students having a quick start in doing NSF-REU physics education research projects. We have duplicated that confirmatory factor analysis example as a Microsoft Excel 2010 example, which is available for download freely from our college website [7].

Conclusions

The sharing of our latent variable model examples with interested faculty would encourage the use of latent variable modeling for assessment beyond the regression method. The LISREL student version working details are discussed above and a Microsoft Excel 2010 confirmatory factor analysis spreadsheet is available for free download on our college website [7]. Future studies could include AI systems for selection among alternative path diagrams [8].

Acknowledgments

Partial supports from several CUNY grants are acknowledged. We thank Alexei Kisselev for laboratory support. We thank Eric Cheung MD, University of Illinois Medical College Psychiatry Department, for discussion. We thank the anonymous reviewers for their suggestions. The free software and open materials are gratefully acknowledged.

References

1 Kevin Dunn, McMaster University, Process Improvement Using Data, free online text, 2017. https://learnche.org/pid/contents
2 Gang Li, S. Joe Qin, Donghua Zhou. A New Method of Dynamic Latent-Variable Modeling for Process Monitoring. IEEE Transactions on Industrial Electronics Volume: 61, Issue: 11, Nov. 2014. http://ieeexplore.ieee.org/document/6718065

3 Dagaev N, Volkova K, Ossadtchi A. Latent variable method for automatic adaptation to background states in motor imagery BCI. J Neural Eng. 2017 Jul 18. https://www.ncbi.nlm.nih.gov/pubmed/28718781
4 Asharf Zulfa Binti Azizan. Structural equation modeling and system dynamic in the detection of power quality event. MSEE Thesis 2012. Universiti Tun Hussein Onn Malaysia Electric and Electronic Engineering. http://eprints.uthm.edu.my/2390/1/ASHRAF_ZULFA.pdf
5 LISREL9.30 student version. http://www.ssicentral.com/lisrel/student.html
6 Jeremy Miles. Confirmatory factor analysis using Microsoft Excel. Behavior Research Methods (2005) 37(4), page 672-676. https://www.researchgate.net/publication/7151927
7 QCC Physics Department website. http://www.qcc.cuny.edu/physics/faculty-research.html http://www.asee.org/papers-and-publications/papers/section-proceedings/middle-atlantic/spring-2012
8 D. Blei. Build, compute, critique, repeat: Data analysis with latent variable models. Annual Review of Statistics and Its Applicaton 1:203-232, 2014. http://www.cs.columbia.edu/~blei/papers/Blei2014b.pdf

Vazgen Shekoyan, PhD

Shekoyan serves as Associate Professor at CUNY Queensborough Community College. His research interests include education research and education material development.

Raul Armendariz, PhD

Armendariz serves as Assistant Professor at CUNY Queensborough Community College. His interest include high energy physics, cosmic ray study and education research

Sunil Dehipawala PhD

Dehipawala serves as Associate Professor at CUNY Queensborough Community College. His research interests include X-ray absorption, random sequence analysis, and education research.

Rex Taibu PhD

Taibu serves as Assistant Professor at CUNY Queensborough Community College. His research interests include education research and education material development.

George Tremberger, BS

Tremberger serves as Lecturer at CUNY Queensborough Community College.

David Lieberman, PhD

Lieberman serves as Professor and Physics Chair at CUNY Queensborough Community College.

Tak Cheung, PhD

Cheung serves as Professor at CUNY Queensborough Community College.