

R을 이용한 PLS 구조방정식모형 분석 튜토리얼: 예시 연구모형 및 데이터를 중심으로

A Tutorial on PLS Structural Equating Modeling using R: (Centering on) Exemplified Research Model and Data

윤 철 호 (Cheolho Yoon) 국립목포대학교 경영학과, 교신저자
김 상 훈 (Sanghoon Kim) 광운대학교 경영학부

요 약

본 튜토리얼은 PLS 경로분석 수행을 지원하는 R 패키지인 “plspm”을 이용하여 구조방정식모형을 분석하는 방법을 제시하고 있다. 본 튜토리얼에서는 이를 위해 기존 연구들에서 제시하고 있는 PLS 구조방정식모형을 분석하기 위한 기준들을 정의하고, 하나의 예시 연구모형을 제시하여 실질적으로 R을 이용하여 이 예시 모형을 분석하는 것을 보여준다. 본 튜토리얼은 PLS를 처음 접하는 연구자들에게는 연구모형에 대한 실증분석을 위한 유용한 가이드가 될 것이며, 이미 PLS에 익숙한 연구자들에게는 R을 이용한 새로운 PLS 구조방정식모형 분석기법 제시를 통하여 R이라는 통합된 통계 소프트웨어 운영환경에서 심도 있는 분석을 위한 기반 지식을 제공할 것이다.

키워드 : R 튜토리얼, 구조방정식모형, Partial Least Squares (PLS), 기술수형모형, plspm, 경로모델링

I. 서 론

최근 경영학을 위시한 제반 사회과학연구에서 PLS(Partial Least Squares)를 이용한 구조방정식 모형(Structural Equation Modeling, SEM)을 근간으로 실증분석을 수행하는 연구들이 급속하게 증가하고 있는 실정이다. PLS 구조방정식모형은 기존의 공분산기반의 구조방정식모형(covariance-based SEM)과 달리 요인기반의 구조방정식모형(component-based SEM)으로 상대적으로 표본크기나 잔차분포에 대한 요구사항에 덜 엄격하여 탐색적인 연구를 수행하거나 복잡한 모형을 분석

하는데 유리한 것으로 알려져 있다(Chin, 1998a). 또한 최근 마케팅을 비롯한 경영정보시스템, 회계 등의 여러 학문영역에서 제기되고 있는 반영지표와 조형지표 사용에 대한 논의(Jarvis *et al.*, 2003, Petter *et al.*, 2007, Rodgers and Guiral)의 확산으로 공분산기반의 구조방정식모형에서는 지원하지 않은 조형지표 모형을 지원한다는 점에서 PLS 구조방정식모형이 더욱 각광을 받고 있다(Monecke and Leisch, 2012). 상기와 같은 이유들로 최근 PLS 구조방정식모형을 근간으로 하는 연구들이 급증함에 따라 이를 지원하기 위한 SmartPLS, PLS-Graph와 같은 PLS 분석 통계 소

프트웨어들도 제 연구들에서 활발하게 사용되고 있고, 이의 적절한 사용절차 및 분석방법에 대한 상세 지침들(e.g., Gefen and Straub, 2005, Wong, 2013)도 학술지를 통해 제시되고 있다.

한편, 최근 빅데이터의 열풍으로 자료들을 분석을 위한 도구로 R이 실무 현장에서 인기를 끌고 있다. R은 통계 프로그래밍 언어이자 오픈 소프트웨어 환경으로 통계, 그래픽, 데이터마이닝 등의 다양하고 방대한 양의 패키지들을 지원한다. R에서 제공되는 패키지들이 오픈 소스이고 선형 및 비선형 모델링, 고전적인 통계분석, 시계열 분석, 분류 및 군집분석 등의 다양한 통계 패키지들을 제공한다는 측면에서 R은 실무는 물론 학문적인 측면에서도, 특히 통계를 기반으로 실증분석을 수행하는 사회과학연구들에서 중요한 역할을 할 수 있을 것으로 기대된다. “plspm”은 R에서 PLS 경로분석(PLS Path Modeling) 수행을 지원하기 위한 패키지로써(Sanchez *et al.*, 2013), R 패키지들을 생산하고 분배하기 위한 저장소인 CRAN(the Comprehensive R Archive Network: <http://cran.r-project.org/web/packages/plspm/index.html>)에서 사용자들은 무료로 획득할 수 있다.

본 튜토리얼은 PLS 경로분석 수행을 지원하는 R 패키지인 “plspm”을 이용하여 구조방정식 모형을 분석하는 방법을 제시하고자 한다. 이를 위해 본 튜토리얼에서는 기존 연구들에서 제시하고 있는 PLS 구조방정식모형을 분석하기 위한 기준들을 정의하고, 하나의 예시 연구모형을 제시하여 실질적으로 R을 이용하여 이 모형을 분석하는 것을 보여준다. 본 튜토리얼은 PLS를 처음 접하는 연구자들에게는 연구모형에 대한 실증분석을 위한 유용한 가이드가 될 것이며, 이미 PLS에 익숙한 연구자들에게는 R을 이용한 새로운 PLS 구조방정식모형 분석기법 제시를 통하여 R이라는 통합된 통계 소프트웨어 운영환경에서 심도 있는 분석을 위한 기반 지식을 제공할 것이다.

II. 이론적 배경

2.1 R을 이용한 통계분석

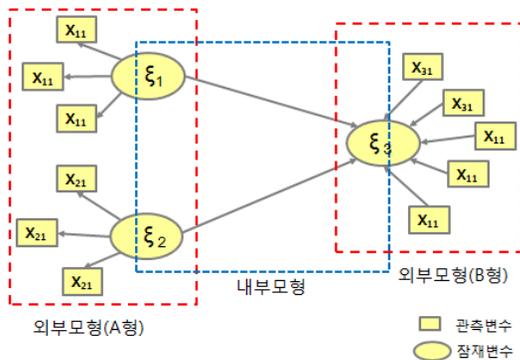
R은 통계 프로그래밍 언어인 S언어 기반으로 만들어진 통계 및 그래픽을 위한 프로그래밍 언어이자 오픈 소프트웨어 환경이다. 뉴질랜드 오클랜드 대학의 Ross Ihaka와 Robert Gentleman에 의해 시작된 R은 통계 소프트웨어 개발과 자료 분석에 광범위하게 사용되고 있으며, 패키지 개발이 용이하여 최근 통계 소프트웨어를 개발하는데 많이 활용된다(Wikipedia, 2014). R로 개발된 패키지들은 GNU GPL(General Public License)하에 무료로 배포되어 일반인들에게 사용되고 있으며, R을 위한 오픈 소스 프로젝트에서는 사용자가 직접 R에 기반한 응용 패키지를 생산하고 자유로이 배포할 수 있도록 지원하기 위하여 패키지 유통 저장소인 CRAN(the Comprehensive R Archive Network)을 운영하고 있다(김광섭, 이기원, 2014). 2014년 11월 현재 CRAN에는 수학, 통계, 그래픽, 데이터마이닝 등을 위하여 6100개 이상의 패키지들이 이용 가능하며(<http://cran.r-project.org> 참조), 특히 최근 빅데이터의 열풍으로 R로 개발된 패키지들이 실무분야에서도 빅데이터를 분석하기 위한 도구들로 적극 활용되고 있다.

R을 이용한 통계분석은 실무분야에서뿐만 아니라 연구 및 교육을 수행함에 있어서도 다음과 같은 장점들을 가지고 있다. 첫째, SPSS나 AMOS와 같은 고가의 통계패키지를 대신하여 R은 무료로 배포되어 연구나 교육을 수행하는데 비용적인 측면에서 유리하다. 둘째, 통계는 물론 수학, 그래픽, 데이터마이닝 등의 다양한 패키지들을 지원하여 통합적인 환경에서 연구와 교육이 가능하다. 셋째, R은 윈도우, 맥 OS 및 리눅스를 포함한 UNIX 운영체제에서 이용이 가능하여 한정된 자원을 보유하고 있는 연구기관이나 교육기관에서 사용하는데 제한이 없다. 넷째, 다양한 최신 통계분석 및 데이터마이닝 기능이 포함된

패키지 및 샘플들이 지속적으로 제공되고 있어 선진적인 교육이나 최신 통계기법을 이용한 연구가 가능하다. 마지막으로 시각화 및 프로그래밍이 가능하여 좀 더 유연하고 창의적인 사고에 기반한 연구나 교육이 가능하다. 상기와 같은 장점들 때문에 R을 이용한 통계분석은 연구나 교육 현장에서 적극적으로 장려될 필요가 있다.

2.2 PLS 구조방정식모형

PLS 구조방정식모형은 <그림 1>과 같이 외부모형(outer model)과 내부모형(inner model)으로 구성되어 있다. 외부모형은 관측변수들(manifest variables)과 그와 관련된 잠재변수(latent variables)와 관계에 대한 모형으로 측정모형(measurement model)이라고도 불리며, 내부모형은 잠재변수들 간의 관계에 대한 모형으로 구조모형(structural model)이라고도 불린다. PLS 구조방정식모형에서 잠재변수는 반영지표 모형(reflective measurement model) 또는 조형지표 모형(formative measurement model)으로 구성될 수 있다. <그림 1>의 외부모형(A형)인 반영지표 모형은 잠재변수가 관측변수에 영향을 미치는 것으로 가정하는 것이며 반영지표 모형에서 잠재변수와 각 측정변수간의 관계는 상관관계이며, 요인분석에서의 요인적재량과 동일하다. <그림 1>의 외부모형(B형)인 조형지표 모형은 관측변수가 잠재변



<그림 1> PLS 구조방정식모형

수로부터 영향을 받는 반영지표와는 달리 잠재변수가 관측변수들에 의해 영향을 받는 것으로 구성된 형태로 관측변수들을 독립변수들로, 잠재변수를 종속변수로 하는 다중회귀분석 모형과 유사하다.

PLS 구조방정식모형 분석은 공분산기반 구조방정식모형 분석 기법들(e.g., Lisrel, AMOS)에 비하여 적은 표본크기를 요구하고, 특히 고객만족도(customer satisfaction)와 같이 일반적으로 비정규 분포를 이루는 자료들에 대한 개념들을 포함하는 연구모형 분석에 유리하여 탐색적인 연구에 적합한 것으로 평가되고 있다(Henseler and Sarstedt, 2013). 또한 최근 사회과학연구들에서 관심과 논의가 증대되고 있는 조형지표 모형을 포함하는 연구모형을 분석할 수 있으며, 그룹비교, 조절변수 분석, 이차요인분석 등에도 유리한 것으로 평가되고 있다. 따라서 최근 많은 사회과학연구들이 상기와 같은 이유들로 PLS 구조방정식모형을 그들 연구들의 분석기법으로 선호하고 있는 실정이다.

PLS 구조방정식모형(또는 PLS 경로분석) 기법은 Wold(1982)에 의하여 개발되었고 Lohmöller(1987)에 의하여 최초의 DOS 기반 소프트웨어인 LVPLS 1.8로 개발되었다. 이후 PLS-GUI(Li, 2005), Visual PLS(Fu, 2006), PLS-Graph(Chin, 2003), XLSTAT-PLSPM(Addinsoft, 2011), SmartPLS(Ringle et al., 2005)와 같은 다양한 윈도우 기반의 PLS 구조방정식모형 분석용 소프트웨어들이 개발되어 사용되고 있으며, 이중 Chin(2013)에 의해 개발된 PLS-Graph와 사용자 인터페이스 기능이 뛰어난 SmartPLS가 최근 연구자들에게 가장 적극적으로 활용되고 있다.

2.3 R 패키지 “plsmpm”

최근 구조방정식모형 분석을 지원하기 위한 R 패키지들이 개발되어 제공되고 있으며, 이중 Sanchez(2013)가 개발한 “plsmpm”은 PLS 경로분석

을 지원하기 위한 R 패키지로 다음과 같은 장점들을 가지고 있다. 첫째, 다른 R 패키지들과 같이 무료 통계소프트웨어이며, 심지어 사용자 등록 및 사용허가 갱신 절차도 필요 없다. PLS-Graph는 비록 무료이기는 하나 사용자 등록 및 매년 사용허가 갱신과 같은 다소 번거로운 절차들을 요구하고 있으며, SmartPLS는 학생버전은 무료이나 전문가 버전은 고가의 요금을 책정하고 있다. 둘째, 기술통계와 구조방정식모형 분석을 R이라는 하나의 통합된 환경에서 수행하는 것이 가능하다. 기존의 PLS 구조방정식모형을 분석하는 연구들을 살펴보면 기술통계는 SPSS나 SAS에서 수행하고 구조방정식모형 분석은 PLS 통계소프트웨어를 사용하는 형태로 진행하지만 R에서는 다양한 패키지의 활용으로 기술통계와 구조방정식모형에 대한 분석을 하나의 통합된 통계분석환경에서 수행할 수 있다. 셋째, 연구분석의 자동화가 가능하다. 즉, 스크립트(script)형태의 언어인 R의 특성을 이용하여 구조방정식모형 분석시 요구되는 신뢰도 평가, 타당도 평가와 같은 절차를 자동화 시킬 수 있으며, 이를 통해 연구분석의 효율성을 높이고 오류발생의 확률을 낮출 수 있다. 넷째, 매개변수 분석은 물론 조절변수분석, 그룹비교, 이차요인분석에 용이하다. PLS-Graph나 SmartPLS를 사용하여 이러한 분석들을 할 경우 엑셀 작업이나 일부의 수작업이 필요하다. 예를 들면 매개변수 분석을 위해 Sobel Test를 수행할 경우 PLS 통계소프트웨어에서 산출된 경로계수와 오차들을 기록한 후 엑셀 또는 분석을 위한 웹사이트를 통해 별도 분석을 해야 한다. 그러나 R환경에서는 Sobel Test와 관련한 수식들을 명령어들로 결합하여 통합처리가 가능하다. 다섯째, 신뢰구간, 직접효과 및 간접효과 등과 같은 기존의 PLS 통계소프트웨어 보다 다양한 정보를 제공한다. 아울러 이러한 정보들은 R의 탁월한 그래픽 기능과 결부되어 보다 시각적으로 다양한 분석을 할 수 있도록 도와준다. 여섯째, 사용자가 R사용에 보다 익숙해질 경우

그래픽 사용자 인터페이스(GUI)를 지원하는 PLS-Graph나 SmartPLS보다 명령어 지향 인터페이스(command-oriented interface)를 지원하는 R 방식이 효율적이고 편리하다(Fox and Andersen 2005). 마지막으로 사용자가 R을 사용할 수 있는 능력수준이 높아짐에 따라 통계분석을 수행함에 있어서 보다 창의적이고 심층적인 분석을 시도할 수 있게 된다.

III. PLS 구조방정식모형 분석시 평가 기준

PLS 구조방정식모형 근간의 실증연구들이 급증함에 따라 최근 마케팅(Hair *et al.*, 2012b)을 비롯하여 경영정보시스템(Urbach and Ahlemann, 2010, Ringle *et al.*, 2012), 전략경영(Hair *et al.*, 2012a), 국제마케팅(Henseler *et al.*, 2009), 운영관리(Peng and Lai, 2012), 회계(Lee *et al.*, 2011) 등의 다양한 경영학 학문영역에서 PLS 구조방정식모형을 이용한 기존 연구들에 대한 비판적 검토(critical review)와 바람직한 PLS 구조방정식모형 분석을 위한 실행지침들이 제시되고 있다. 이러한 연구들에서 제시하는 PLS 구조방정식모형 분석을 위한 외부모형 평가와 내부모형 평가에 대한 기준들을 다음과 같다.

3.1 외부모형 평가 기준

외부모형(또는 측정지표모형)에 대한 평가는 관측변수들 즉, 측정지표들에 대한 신뢰도 및 타당도에 대한 평가로서 PLS 구조방정식모형에 있어서 외부모형의 평가는 반영지표 모형과 조형지표 모형에 따라 측정지표들에 대한 신뢰도 및 타당도 평가방식 및 기준들이 달리 설정되어야 한다.

3.1.1 반영지표 모형

PLS 구조방정식모형에서 반영지표 모형의

측정지표들에 대한 신뢰도와 타당도의 평가기준은 다음과 같다. 우선 신뢰도 평가는 기존의 연구들에서 3가지의 기준들이 제시되고 있다. 첫째, 전통적인 신뢰도 평가방식인 내적 일관성 신뢰도(internal consistency reliability) 평가방식으로 클론바흐 알파(Cronbach's alpha) 계수를 평가 항목으로 하고, 신뢰도 확보를 위한 기준값은 0.7 이상으로 제시되고 있다. 둘째 기준은 합성신뢰도(composite reliability)로서 이 역시 내적 일관성 평가에 기반한 신뢰도이나 클론바흐 알파계수가 모든 측정지표들이 동등하게 신뢰할 수 있다는 가정하에 산출됨에 반하여 합성신뢰도는 아래 수식과 같이 측정지표들이 서로 다른 적재량임을 고려하여 산출됨으로써 구조방정식모형 분석에서 클론바흐 알파계수 보다 바람직한 신뢰도 평가 항목으로 평가되고 있다(Henseler *et al.*, 2009). 기존 연구들에서 합성신뢰도 확보를 위해 일반적으로 제시되고 있는 기준값은 Bagozzi and Yi (1988)의 연구에서 제시된 0.7 이상이다.

$$\text{합성신뢰도(CR)} = (\sum \lambda)^2 / ((\sum \lambda)^2 + \sum \text{Var}(\varepsilon))$$

λ : 표준화된 적재량, $\text{Var}(\varepsilon)$: 측정오차

셋째 기준은 지표 신뢰도(indicator reliability)로서 지표 신뢰도는 측정지표들이 측정하고자 하는 개념에 대하여 일관성이 있는지에 대한 정도로 정의되며, 지표 신뢰도를 확보하기 위해서는 측정지표들의 표준화된 적재량들(standardized loadings)이 0.7을 상회해야 한다. 그러나 탐색적인 연구인 경우 연구자들의 자료조사 상황에 대한 합리적 판단에 따라 0.6 이상도 허용될 수 있다(Hair *et al.*, 2012b).

PLS 구조방정식모형분석에서 반영지표모형의 관측변수들, 즉 측정지표들에 대한 타당도 평가는 집중타당도(Convergent validity) 및 판별타당도(Discriminant validity) 평가결과에 의거한다. 기존연구들에서 PLS 구조방정식모형 분석에서의 측정지표들에 대한 집중타당도 평가 기준은

두 가지로 제시되고 있다. 첫째, 측정지표들의 잠재변수들에 대한 적재값들의 유의성에 대한 평가로서 측정지표들의 표준화된 적재량이 통계적으로 유의한 수준($t\text{-value} > 1.96$)이면 집중타당도가 확보되었다고 본다. 둘째, 잠재변수들의 평균분산추출(Average Variance Extracted, AVE) 값이며 Fornell and Larcker(1981)에 의해 제시된 평균분산추출(AVE)은 잠재변수에 대해 측정지표들이 설명할 수 있는 분산의 크기로 아래와 같은 식으로 계산되며, 이 결과값이 0.5 이상이면 집중타당도가 확보된 것으로 본다.

$$\text{평균분산추출(AVE)} = \sum \lambda^2 / (\sum \lambda^2 + \sum \text{Var}(\varepsilon))$$

λ : 표준화된 적재량, $\text{Var}(\varepsilon)$: 측정오차

기존 연구들에서 제시된 측정지표들로 구성된 잠재변수들 대한 판별타당도에 대한 평가기준은 세 가지로 요약될 수 있다. 첫째, Fornell-Larcker 기준이며, 이 기준은 한 잠재변수의 평균분산추출(AVE) 값이 그 잠재변수와 다른 잠재변수들 간의 상관계수의 제곱 값들보다 높을 때 판별타당도가 확보되었다고 보는 것이다(Fornell and Larcker, 1981). 둘째로는 교차적재량(Cross loadings) 기준인데 이는 확인적 요인 분석 결과, 각 잠재변수들에 속한 측정지표들의 적재량들이 다른 잠재변수들에서의 교차적재량에 비해 뚜렷하게 높게 적재되고 있어야 판별타당성이 있다고 보는 것이다(Chin, 1998b). 셋째로는 Gefen and Straub(2005)가 제시한 기준이며, 이 기준은 각 잠재변수의 평균분산추출(AVE) 값의 제곱근이 해당 잠재변수와 다른 잠재변수들 간의 상관계수 값들 보다 높을 때 판별타당도가 확보되었다고 보는 것인데 이 기준은 첫 번째 기준으로 제시된 Fornell-Larcker 기준의 변형으로서 같은 결과를 나타내면서 보다 산출과정이 편리하여 PLS 구조방정식모형 분석에서 Gefen and Straub(2005)의 기준을 사용하는 것이 바람직해 보인다.

3.1.2 조형지표모형

PLS 구조방정식모형분석에서 조형지표 모형의 들에 대한 신뢰도 및 타당도 평가기준으로서 세 가지 기준이 제시되고 있다. 이들 중 두 가지 기준은 지표타당도 확보를 위한 평가기준으로서 첫째, 각 잠재변수에 대한 측정지표들의 가중치들(weights)의 통계적 유의성수준 평가를 위한¹⁾ t-값이 최소 1.96 이상이 되어야 한다고 보고 있으며, 둘째 측정지표들간의 다중공선성(Multicollinearity)에 대한 평가로서 분산팽창계수(Variance Inflation Factor, VIF)가 일반적인 회귀분석에서 요구하는 10 이하보다 엄격한 수치인 3.3 이하이어야 한다고 주장되고 있다(Diamantopoulos and Sigauw, 2006). 그러나 다수의 연구들은 5.0 이하를 제시하고 있다(Hair et al., 2011b). 세 번째 기준은 개념 타당도를 평가하기 위한 기준이며 조형지표 모형의 경우 집중타당도 평가는 의미가 없으며, 판별타당도 평가는 반영모형과 같은 방식들로 평가되는 것인 바람직한 것으로 주장되고 있다(MacKenzie et al., 2005). 아울러 조형지표 모형에 대한 타당도 평가시에는 비록 측정지표들에 대한 통계적 분석에 의한 평가는 아니지만 관련 연구분야에 대한 충분한 문헌 고찰을 통한 이해타당도(Nomological validity) 검증의 중요성이 강조되고 있다(Henseler et al., 2009).

3.2 내부모형 평가 기준

내부모형에 대한 평가는 잠재변수들간의 관계 즉, 구조모형에 대한 평가이다. 내부모형에 평가는 경로계수에 대한 평가, 결정계수 R²에 대한 분석 및 모형적합도에 대한 평가 등이 포

함된다. 경로계수(path coefficients)에 대한 평가는 잠재변수들간의 인과관계에 대한 평가로서 부호, 크기나 중요성이 고려되어 평가되어야만 하는 것으로 주장(Hair et al., 2012b)되고 있지만, 실제 대부분의 연구들에서는 경로계수의 통계적 유의성수준(p < 0.05) 위주로 평가되고 있다. 즉, PLS 구조방정식모형분석에서 경로계수들의 유의성은 부트스트랩(bootstrap) 샘플링 방식을 통한 경로계수 추정치와 변량(오차)에 근간한 t-값을 이용하여 평가되는데 최근 연구들에 따르면 부트스트랩은 비모수(non-parametric)를 평가하기 위한 방법이기 때문에 신뢰구간(confidence intervals)에 의한 평가 방식이 바람직하다는 견해가 제시되고 있다(Henseler et al., 2009). 결정계수 R²에 대한 평가는 내생 잠재변수들(endogenous latent variables)에 대한 설명력으로 명확한 기준 값들은 없으나 Chin(1998b)은 0.19은 약함, 0.33는 중간, 0.67 이상은 설명력이 큰 것으로 제안하고 있다. 모형의 적합도에 대한 평가는 PLS 구조방정식모형 분석이 기존의 공분산기반의 구조방정식모형 분석에 비하여 가장 취약한 부분으로 Tenenhaus et al. (2005)는 공통성(communality) 값과 R²를 이용하여 GoF(Goodness of Fit) 제시하고 기준값으로 0.36 이상을 권고하고 있으나, 아직 이 기준은 연구자들에 의해 폭넓은 지지를 받고 있지는 못하고 있다(Henseler and Sarstedt, 2013). 따라서 PLS 구조방정식모형 분석에 있어서 적합도 수용에 대한 절대적인 기준은 설정할 필요는 없다고 본다.

3.3 PLS 구조방정식모형 분석기준 정립

기존의 PLS 구조방정식모형 분석을 위한 지침에 관한 연구들에서 제시하는 기준들에 근거하여 본 연구에서 통합적으로 정립한 외부모형 평가와 내부모형 평가에 대한 기준들을 제시하면 다음의 <표 1>과 같다.

1) 조형지표 모형에서는 잠재변수와 각 측정변수간의 상관계수로 산출한 표준화된 적재량이 아니라 잠재변수와 각 측정변수간의 경로계수와 유사한 가중치들(weights)이 의미가 있다.

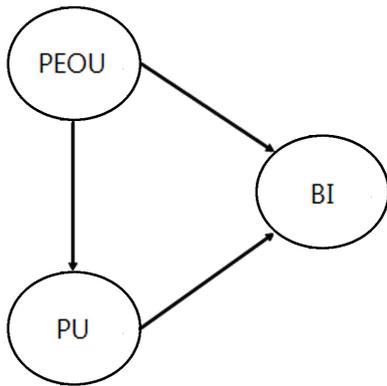
〈표 1〉 PLS 구조방정식모형 분석기준

외부모형 평가(반영지표 모형)		
신뢰도		
내적 일관성 신뢰도	클론바흐 알파계수 ≥ 0.70	Nunnally and Bernstein (1994)
	합성신뢰도 ≥ 0.70	Bagozzi and Yi(1988)
지표 신뢰도	측정지표들의 요인 적재값들 ≥ 0.70 (탐색적 연구는 0.6도 수용가능)	Chin(1998b) Hair <i>et al.</i> (2012b)
집중타당도		
측정지표들의 유의도	t-값 > 1.96 (부트스트랩을 사용하여 추정)	Gefen and Straub(2005)
표준분산추출(AVE)	표준분산추출(AVE) ≥ 0.50	Fornell and Larcker(1981)
관별타당도		
교차적재기준	확인적 요인분석에서 측정지표들이 다른 잠재변수들보다 그들이 측정하는 잠재변수에 보다 뚜렷하게 높게 적재되고 있어야 함	Chin(1998b) Gefen and Straub(2005)
Gefen-Straub의 기준	평균분산추출(AVE)의 제곱근 값이 그 잠재변수와 다른 잠재변수들 간의 상관계수 값들 보다 높아야 함	Gefen and Straub(2005)
외부모형(조형지표 모형)		
타당도		
지표 타당도	가중치의 유의도 t-값 > 1.96	Chin(1998b)
	측정지표들 간의 다중공선성 점검 VIF < 5.0	Hair <i>et al.</i> (2011b)
개념 타당도	반영지표모형과 같은 방식의 관별타당도 평가 (Gefen-Straub의 기준)	Mackenzie <i>et al.</i> (2005)
내부모형 평가		
경로분석		
경로계수	$p < 0.05$ (부트스트랩을 사용하고, 신뢰구간이 제공되는 것이 바람직함)	Henseler <i>et al.</i> (2009)
모형의 설명력		
결정계수(R^2)	절대적인 기준 없음, 일반적으로 일 0.19은 약함, 0.33는 중간, 0.67 이상은 큰 것으로 알려져 있음	Chin(1998b)
모형 적합도		
GOF(Goodness of Fit)	권고기준은 GoF > 0.36 이나 절대적인 수용 기준은 없음	Tenchaus <i>et al.</i> (2005)

IV. R을 이용한 PLS 구조방정식모형 분석 예시

R을 이용한 PLS 구조방정식모형 분석절차 및 방법을 구체적으로 보여주기 위하여, 경영 정보시스템 분야 연구에서 가장 잘 알려진 이론적 모형들 중 하나인 기술수용모형(Technology Acceptance Model, TAM) (Davis 1989)을 사

용한 연구모형을 제시하고 본 모형과 관련하여 수집된 데이터를 중심으로 실증 분석하는 과정을 보이고자 한다. 즉, 분석대상 연구모형은 TAM에서와 같이 지각된 사용용이성(PEOU)과 지각된 유용성(PU)이 행위의도(BI)에 영향을 주고, 더불어 지각된 사용용이성(PEOU)이 지각된 유용성(PU)가 영향을 주는 것으로 모형화하였다.



〈그림 2〉 예시를 위한 연구모형

또한 예시 연구모형 검증을 위한 데이터는 기존의 연구(Yoon and Kim, 2007)에서 사용된 데이터의 일부가 PLS 실행을 통한 연구모형 분석 과정을 보여 주기 위하여 사용되었다. 분석에 사용된 측정지표들은 <표 2>와 같으며, 데이터는 리커트 7점 척도로 수집되었고 <부록 1>에 첨부되어 있다.

4.1 R에서의 구조방정식모형 구성과 PLS 실행

R에서 PLS을 실행하기 위해서는 우선 “plspm”

패키지가 설치되고 라이브러리로 적재되어야만 한다. 이를 위한 R에서의 명령어는 다음과 같다.

```

# plspm 패키지 설치
> install.packages("plspm")

# plspm 라이브러리 로드
> library(plspm)
  
```

“plspm” 패키지가 설치가 된 후, PLS 분석을 위하여 설문자료를 R로 불러들이고 자료를 확인한다. 이를 위한 명령어들은 아래와 같다. <그림 3>은 R로 불러 들어온 자료를 보여주고 있다. 참고로 자료(TAM.csv)는 작업영역 디렉터리(예: C:/PLS_PM)에 보관되어 있으며 작업영역을 설정하는 명령어가 미리 실행되어야 한다.

```

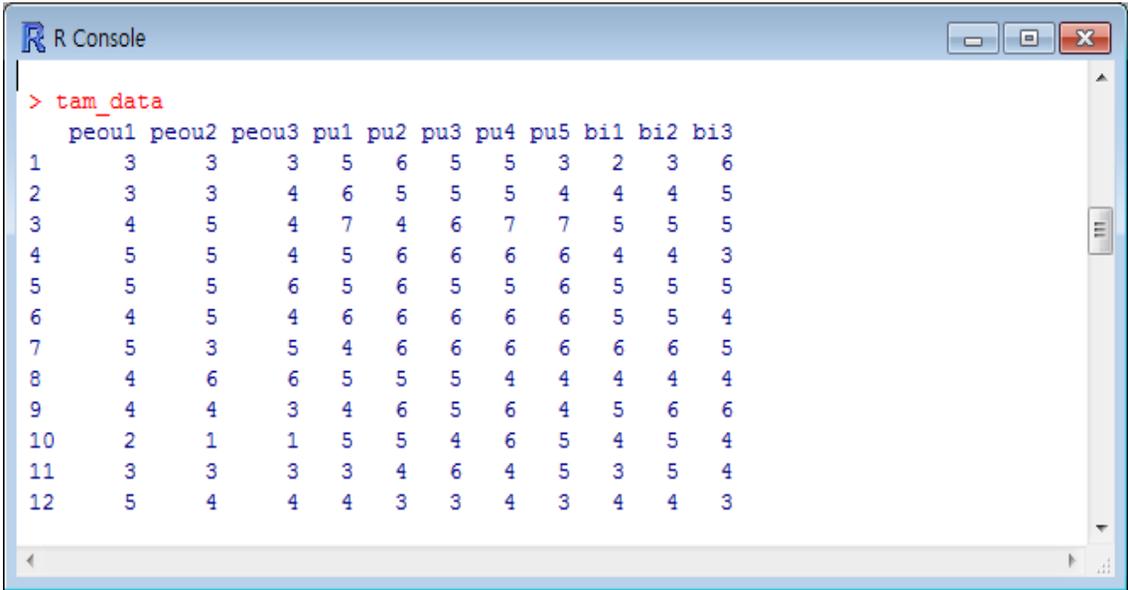
# 작업영역 설정
> setwd("C:/PLS_PM")

# 설문자료 R로 로드
> tam_data <- read.csv("TAM.csv", header = T)

# 설문자료 보기
> tam_data
  
```

〈표 2〉 측정지표

개념	코드	측정지표
PEOU	peou1	나는 일을 하는데 무선랜을 사용하는 것이 쉽다는 것을 발견한다.
	peou2	무선랜을 익숙하게 사용하는 것은 쉬운 일이다.
	peou2	나는 무선랜이 사용하기에 쉽다는 것을 발견한다.
PU	pu1	무선랜을 이용하면 내 업무를 더욱 신속하게 처리할 수 있다.
	pu2	무선랜을 이용하면 내 업무성과를 향상시킬 수 있다.
	pu3	무선랜을 이용하면 내 업무생산성을 증대시킬 수 있다.
	pu4	무선랜을 이용하여 내 업무를 하는 것이 효과적이다.
	pu5	나는 무선랜이 내 업무에 유용하다는 것을 발견한다.
BI	bi1	나는 무선랜을 사용하는 것을 즐길 것이다.
	bi2	나는 무선랜을 빈번하게 사용할 것이다.
	bi3	나는 다른 사람들에게 무선랜을 사용하라고 권유할 것이다.



```

> tam_data
  peou1 peou2 peou3 pu1 pu2 pu3 pu4 pu5 bi1 bi2 bi3
1      3      3      3  5  6  5  5  3  2  3  6
2      3      3      4  6  5  5  5  4  4  4  5
3      4      5      4  7  4  6  7  7  5  5  5
4      5      5      4  5  6  6  6  6  4  4  3
5      5      5      6  5  6  5  5  6  5  5  5
6      4      5      4  6  6  6  6  6  5  5  4
7      5      3      5  4  6  6  6  6  6  6  5
8      4      6      6  5  5  5  4  4  4  4  4
9      4      4      3  4  6  5  6  4  5  6  6
10     2      1      1  5  5  4  6  5  4  5  4
11     3      3      3  3  4  6  4  5  3  5  4
12     5      4      4  4  3  3  4  3  4  4  3

```

〈그림 3〉 R로 불러온 설문자료

설문자료를 R로 불러들여 이 자료에 대한 PLS 분석을 하기 위해서는 1) 연구모형의 잠재변수들간의 경로를 행렬 구조로 변경하고, 2) 잠재변수와 관련된 관측변수(측정지표)들에 대하여 정의하고, 3) 관측변수들의 형태를 정의(반영지표 vs. 조형지표)하며, 마지막으로 4) PLS을 실행하는 단계로 진행된다. PLS 구조방정식모형을 분석하는 경우 내부모형과 외부모형을 순환·반복하며 관측변수(측정지표)들의 가중치(weights)를 산출하는 PLS 알고리즘의 구조상 공분산기반의 구조방정식모형(측정모형과 구조모형 설정의 2 단계 접근방법)과 다르게 이론적 구조모형(잠재변수들 간의 경로)과 측정모형을 동시에 설정하여야만 한다.

〈그림 2〉와 같은 연구모형의 잠재변수들간의 경로를 행렬 구조로 변경하기 위해서는 내부모형의 열 값들을 정의하고 열 결합을 통한 경로행렬을 생성해야 하며 이를 위한 명령어들은 다음과 같다. 내부모형 경로행렬의 열은 0과 1로 구성하며, 선행되는 경로가 있는 경우 1을 없는 경우 0으로 설정한다.

```
# 내부모형 행렬의 열
```

```
> PEOU = c(0, 0, 0)
```

```
> PU = c(1, 0, 0)
```

```
> BI = c(1, 1, 0)
```

```
# 열 결합을 통한 경로행렬 벡터 생성
```

```
> tam_path = rbind(PEOU, PU, BI)
```

```
# 행 이름 추가(선택적)
```

```
> colnames(tam_path) = rownames(tam_path)
```

생성된 내부모형행렬은 다음과 같은 명령어들로 확인하고(〈그림 4〉 참조), 그림으로도 경로를 볼 수 있다.

```
# 경로행렬 보기
```

```
> tam_path
```

```
# 그림으로 경로보기
```

```
> innerplot(tam_path)
```



〈그림 4〉 경로행렬

경로행렬(tam_path)을 생성하고 난 후 다음절차는 잠재변수와 관련된 측정지표들 즉, 관측변수들에 대하여 정의하는 것이다. <그림 3>의 설문자료에서 보이는 바와 같이 첫 번째부터 세 번째 행까지(peou1~peou3)는 지각된 사용용이성(PEOU)에 대한 측정지표들이고 네 번째부터 여덟 번째까지(pu1~pu5)는 지각된 유용성(PU) 그리고 아홉째부터 열한 번째까지(bi1~bi3)는 행위의도(BI)와 관련된 지표들인데 다음과 같은 명령어로 잠재변수와 관련된 관측변수들을 정의한다.

```

# 잠재변수와 관련된 관측변수들 정의
> tam_blocks = list(1:3, 4:8, 9:11)
    
```

다음 절차는 관측변수들의 형태를 정의하는 단계로 반영지표 모형은 “A”로 조형지표 모형은 “B”로 정의한다. 예시 설문문항이 근거한 연구 모형에서는 3개의 모든 잠재변수들이 반영지표들로 측정되었으므로 잠재변수에 대한 관측변수들의 형태를 정의하기 위해 아래와 같은 명령어를 실행한다,

```

# 모든 잠재변수가 반영지표로 측정 되었을 때
> tam_modes = c("A", "A", "A")
    
```

참고로 연구 모형 내 잠재변수들 중 지각된 사용용이성(PEOU)이 조형지표들로 측정 되었다고 가정하면 잠재변수에 대한 관측변수들의 형태를 정의하기 위한 명령어는 다음과 같다.

```

> tam_modes = c("B", "A", "A")
    
```

R에서의 구조방정식모형 구성 및 PLS 실행에서의 마지막 절차는 분석대상 연구모형에 PLS를 실행하는 단계이다. 설문자료(tam_data)와 이전에 정의한 경로행렬(tam_path), 잠재변수와 관련된 관측변수들에 대한 정의(tam_blocks) 및 관측변수들의 형태(tam_modes) 등을 입력 값들로 하여 아래와 같이 PLS를 실행한다.

```

# pls 실행
> tam_pls = plsspm(tam_data, tam_path, tam_blocks,
  modes = tam_modes, boot.val = TRUE, br = 500)
    
```

PLS 실행에서 샘플링 방식은 PLS 구조방정식 모형에서 집중타당도와 경로계수 평가의 기준으로 제시된 부트스트랩을 사용하고(boot.val = TRUE) 샘플링 횟수는 500회 이상으로(br = 500) 하는 것이 바람직한 것으로 알려져 있다(Chin 1998).

4.2 외부모형의 평가

PLS 실행을 통한 외부모형에 대한 평가는 예시 연구모형의 경우 모든 잠재변수들의 측정지표들이 반영지표들로 구성되어 있어 반영지표 모형에 대한 신뢰도, 집중타당도 그리고 판별타당도 평가 순으로 진행된다.

4.2.1 신뢰도 평가

PLS을 이용한 구조방정식모형 분석에서 내적 일관성 신뢰도 평가는 클론바흐 알파계수와 합성신뢰도를 기준으로 평가되며 “plspm”에서는 이와 관련하여 관측변수(측정지표)들에 대한 단일차원성(uni dimension) 정보를 제공한다. 이를 실행하기 위한 명령어는 다음과 같으며 그 결과는 <그림 5>와 같이 나타났다.

```
> tam_pls$unidim
```

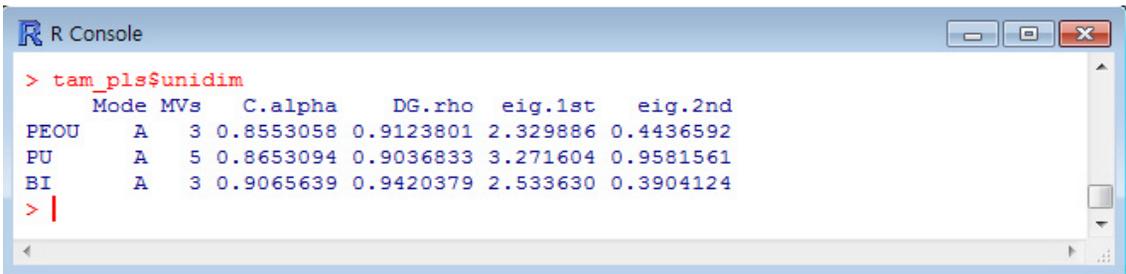
<그림 5>에 보이는 바와 같이 “tam_pls\$unidim”을 실행했을 때, 잠재변수 별로 클론바흐 알파 값(C. Alpha)과 Dillon-Goldstein’s rho값, 그리고 고유치(eigen value)들(eig.1st, eig.2nd)에 대한

정보가 산출된다.

<그림 5>에 보이는 바와 같이 모든 잠재변수들은 클론바흐 알파 값이 0.86 이상으로 기준치인 0.7을 초과하고 합성신뢰도의 다른 이름인 Dillon-Goldstein’s rho 값도 모두 9.0을 넘어 합성신뢰도 기준치인 0.7보다 큰 것으로 나타나 측정지표들의 내적 일관성 신뢰도가 확보됨을 보여주고 있다.

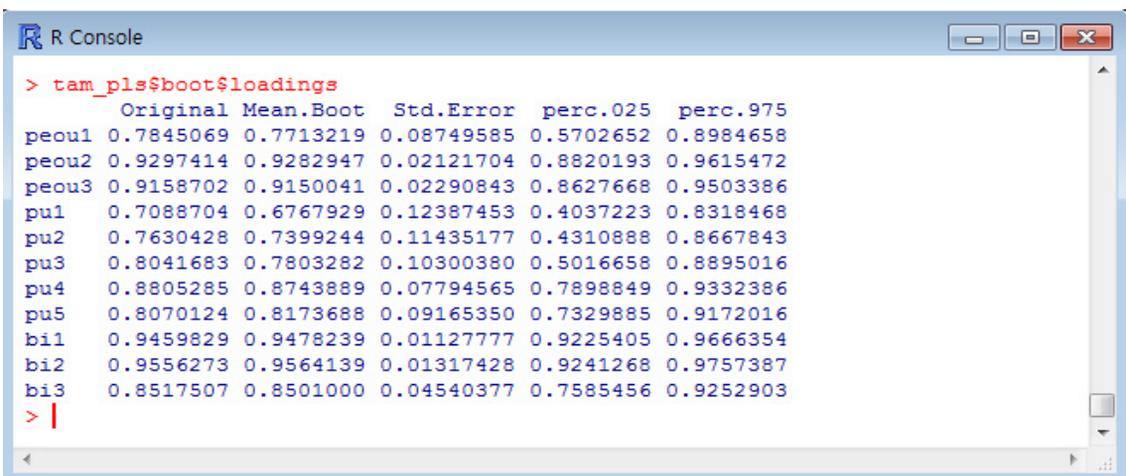
지표 신뢰도는 측정지표들의 표준화된 적재값들이 0.7을 상회해야 하는데 적재값들을 확인하기 위한 명령어는 다음과 같으며, 그 결과는 <그림 6>과 같다.

```
> tam_pls$boot$loadings
```



```
> tam_pls$unidim
      Mode MVs  C.alpha  DG.rho  eig.1st  eig.2nd
PEOU   A    3 0.8553058 0.9123801 2.329886 0.4436592
PU     A    5 0.8653094 0.9036833 3.271604 0.9581561
BI     A    3 0.9065639 0.9420379 2.533630 0.3904124
> |
```

<그림 5> 내적 일관성 신뢰도 평가



```
> tam_pls$boot$loadings
      Original Mean.Boot Std.Error perc.025 perc.975
peou1 0.7845069 0.7713219 0.08749585 0.5702652 0.8984658
peou2 0.9297414 0.9282947 0.02121704 0.8820193 0.9615472
peou3 0.9158702 0.9150041 0.02290843 0.8627668 0.9503386
pu1    0.7088704 0.6767929 0.12387453 0.4037223 0.8318468
pu2    0.7630428 0.7399244 0.11435177 0.4310888 0.8667843
pu3    0.8041683 0.7803282 0.10300380 0.5016658 0.8895016
pu4    0.8805285 0.8743889 0.07794565 0.7898849 0.9332386
pu5    0.8070124 0.8173688 0.09165350 0.7329885 0.9172016
bi1    0.9459829 0.9478239 0.01127777 0.9225405 0.9666354
bi2    0.9556273 0.9564139 0.01317428 0.9241268 0.9757387
bi3    0.8517507 0.8501000 0.04540377 0.7585456 0.9252903
> |
```

<그림 6> 지표 신뢰도 평가

<그림 6>에 보이는 바와 같이 모든 관측변수들의 적재값들(“Original” 수치값)이 기준치인 0.7보다 큰 것으로 나타나 측정지표들의 지표 신뢰도가 확보됨을 보여 주고 있다.

4.2.2 집중타당도 평가

PLS 구조방정식모형에서의 집중타당도의 평가에 대한 기준은 측정지표들의 잠재변수에 속한 적재값들의 통계적 유의성에 대한 평가와 잠재변수들의 평균분산추출(AVE) 값이다. 부트스트랩에 의한 추정치에 대한 통계적 유의성에 대한

검증은 신뢰구간(confidence interval)에 기반한 검증으로 수행되는 것이 논리적이다. 따라서 “plsrm”에서 부트스트랩으로 수행한 경우 <그림 6>과 같이 경험적 분포²⁾ 백분율에 대한 95% 신뢰구간 정보인 최소값(perc.025)와 최대값(perc.975)를 포함하고 있으나, 정규분포를 가정하여 산출되는 t-통계량은 제공하지 않는다. 그러나 기존의 연구들에서는 t-값에 의거한 기준을 제시하고 있으므로 R에서 Chin(1998b)의 방식과 같이 추정치를 오차로 나누는 다음과 같은 명령을 통해 t-값을 생성·추가하며, 그 결과는 <그림 7>과 같다.

```

R Console
> tam_pls$boot$loadings
      Original Mean.Boot Std.Error perc.025 perc.975 t_value
peou1 0.7845069 0.7713219 0.08749585 0.5702652 0.8984658 8.966218
peou2 0.9297414 0.9282947 0.02121704 0.8820193 0.9615472 43.820500
peou3 0.9158702 0.9150041 0.02290843 0.8627668 0.9503386 39.979611
pu1    0.7088704 0.6767929 0.12387453 0.4037223 0.8318468 5.722488
pu2    0.7630428 0.7399244 0.11435177 0.4310888 0.8667843 6.672767
pu3    0.8041683 0.7803282 0.10300380 0.5016658 0.8895016 7.807170
pu4    0.8805285 0.8743889 0.07794565 0.7898849 0.9332386 11.296698
pu5    0.8070124 0.8173688 0.09165350 0.7329885 0.9172016 8.805036
bi1    0.9459829 0.9478239 0.01127777 0.9225405 0.9666354 83.880305
bi2    0.9556273 0.9564139 0.01317428 0.9241268 0.9757387 72.537352
bi3    0.8517507 0.8501000 0.04540377 0.7585456 0.9252903 18.759472
> |
    
```

<그림 7> 추정치에 t-값 추가

```

R Console
> tam_pls$inner_summary
      Type      R2 Block_Community Mean_Redundancy      AVE
PEOU Exogenous 0.0000000      0.7728962      0.0000000 0.7728962
PU   Endogenous 0.1193619      0.6316035      0.07538943 0.6316035
BI   Endogenous 0.4224898      0.8445288      0.35680480 0.8445288
> |
    
```

<그림 8> 평균분산추출(AVE) 값

2) 부트스트랩에서 k회(e.g. 500회) 샘플링 한 자료들을 실행하여 얻은 추정치들로 이루어진 분포.

```
> tam_pls$boot$loadings$t_value = with(tam_pls$
boot$loadings, Original/Std.Error)
> tam_pls$boot$loadings
```

<그림 7>의 오른쪽에 보이는 바와 같이 모든 측정지표들의 적재값들의 t-값이 기준치인 1.96을 초과하고 있음을 보여 주고 있다.

PLS 구조방정식모형에서의 집중타당도의 두 번째 기준은 평균분산추출(AVE) 값으로 잠재변수들의 평균분산추출(AVE) 값을 보기 위해서는 다음과 같은 명령어를 실행하며 그 결과는 <그림 8>과 같다.

```
> tam_pls$inner_summary
```

<그림 8>에서 나타난 바와 같이 모든 잠재변수들의 평균분산추출(AVE) 값이 기준치인 0.5를 상회하고 있음을 보여 주고 있다. 즉, 상기 두 기준 모두를 충족하고 있으므로 측정지표들의 집중타당도가 확보되고 있음을 알 수 있다.

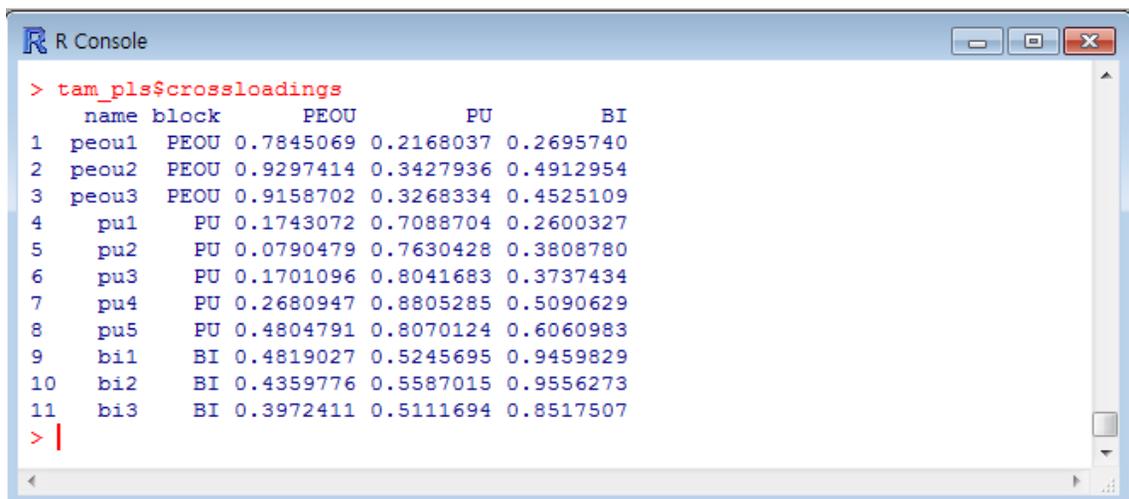
4.2.3 판별타당도 평가

이전 장에서 PLS 구조방정식모형에서의 판별

타당도 평가에 대한 기준으로 교차적재량(Cross loadings)기준과 Fornell-Larcker 기준을 변형한 Gefen and Straub(2005) 기준이 제시되었다. 즉, 확인적 요인분석결과 각 잠재변수별로 속해 있는 측정지표들의 적재량이 다른 잠재변수들에서의 이들 측정지표들의 교차적재량에 비해 뚜렷하게 높게 적재되고 있고, 각 잠재변수의 평균분산추출(AVE) 제곱근 값이 해당 잠재변수와 다른 잠재변수들 간의 상관관계수 값들보다 높을 때 판별타당도가 있다고 보는데, 우선 교차적재량을 보기 위한 명령어는 다음과 같으며, 그 결과는 <그림 9>와 같다.

```
> tam_pls$crossloadings
```

<그림 9>에 나타난 바와 같이 모든 측정지표들의 경우 이들이 속해 있는 측정대상 잠재변수에서의 적재량이 다른 잠재변수에서의 교차적재량 보다 현저히 큼을 보여주고 있다. 즉, PEOU는 측정지표 peou1~peou3에서 높은 수치, PU는 측정지표 pu1~pu5에서 높은 수치 그리고 BI는 측정지표 bi1~bi3에서 높은 수치를 보여 주고 있어 판별타당도 확보를 위한 첫째 기준을 충족하



```
> tam_pls$crossloadings
  name block      PEOU      PU      BI
1 peou1 PEOU 0.7845069 0.2168037 0.2695740
2 peou2 PEOU 0.9297414 0.3427936 0.4912954
3 peou3 PEOU 0.9158702 0.3268334 0.4525109
4 pu1    PU 0.1743072 0.7088704 0.2600327
5 pu2    PU 0.0790479 0.7630428 0.3808780
6 pu3    PU 0.1701096 0.8041683 0.3737434
7 pu4    PU 0.2680947 0.8805285 0.5090629
8 pu5    PU 0.4804791 0.8070124 0.6060983
9 bi1    BI 0.4819027 0.5245695 0.9459829
10 bi2   BI 0.4359776 0.5587015 0.9556273
11 bi3   BI 0.3972411 0.5111694 0.8517507
> |
```

<그림 9> 교차적재량(Cross Loadings)

고 있음을 알 수 있다.

판별타당도의 평가의 두 번째 기준은 한 잠재변수의 평균분산추출(AVE) 제공근 값이 그 잠재변수와 다른 잠재변수들 간의 상관계수 값들보다 높아야 하는 것인데 plspm에서는 잠재변수들간의 상관관계를 제시하는 기본정보가 없다. 따라서 이를 위해서는 Gefen and Straub(2005)의 연구에서 사용한 방법과 같이 잠재변수의 점수값들(latent variable scores)를 이용한 상관관계분석을 통하여 해결해야만 한다. 우선 잠재변수들의 상관계수 값 행렬을 생성하기 위하여 행렬로 되어 있는 잠재변수의 점수값들(tam_pls\$scores)을 데이터프레임 형식으로 자료를 변형 후 잠재변수들간의 피어슨(pearson) 상관계수를 구하며, R에서 이와 관련한 명령어들은 다음과 같다.

```
# 행렬(매트릭스) 데이터프레임으로 변환해 tam_scores 변수에 저장
> tam_scores <- as.data.frame(tam_pls$scores)
# 잠재변수들간의 상관계수 생성
> tam_cor = cor(tam_scores, use = "complete.obs",
method = "pearson")
```

참고로 자료가 분포 가정을 위반하거나 서열척도로 구성되어 있다면 아래의 명령어와 같이 비모수 방식인 스피어만(spearman) 상관관계를 사용하는 것이 바람직하다(Gefen and Straub, 2005).

```
> tam_cor = cor(tam_scores, use = "complete.obs",
```

```
method="spearman")
```

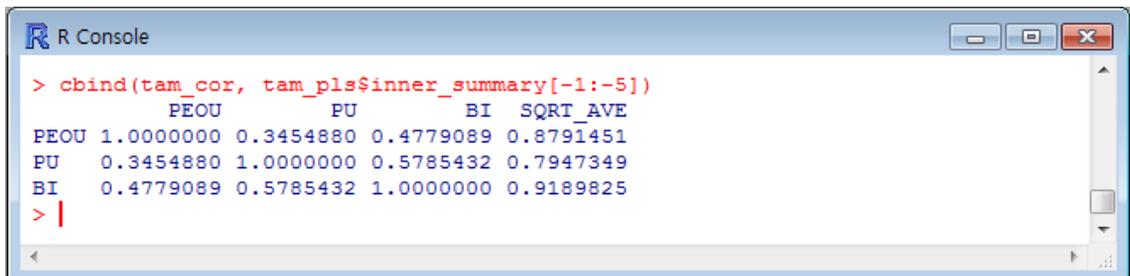
또한 생성한 잠재변수들 간의 상관계수 행렬에 잠재변수의 평균분산추출(AVE) 제공근 값들은 다음과 같은 명령어로 생성·추가 가능하며, 이를 통해 <그림 10>과 같이 잠재변수들간의 상관계수 행렬과 각 잠재변수의 제공근 값을 하나의 테이블로 산출할 수 있다.

```
> tam_pls$inner_summary$SQRT_AVE = with(tam_pls$inner_summary, sqrt(AVE))
> cbind(tam_cor, tam_pls$inner_summary[-1:-5])
```

<그림 10>에서 나타난 결과는 모든 잠재변수들의 평균분산추출 제공근 값들이 그 변수와 다른 변수들간의 상관계수 값보다 높은 것으로 보여주고 있어 판별타당도 확보의 두 번째 기준도 충족됨을 알 수 있으며, 이에 따라 예시 데이터에 의거할 경우 개념들간의 판별타당도가 있음을 확인할 수 있다.

4.2.4 조형지표 타당도 평가

비록 본 연구모형에서 사용된 측정지표들이 조형지표들은 아니나, 본 튜토리얼에서는 제시한 개념중 지각된 사용용이성(PEOU)이 조형지표들로 구성되었다고 가정하고 이에 대한 타당도 분석기준에 대한 검증방법을 보여주고자 한다. 우선 관측변수들의 형태를 아래와 같이 정의하고 PLS를 실행한다.



<그림 10> 상관계수와 평균분산추출의 제공근값

```
> tam_modes = c("B", "A", "A")
> tam_pls = plsprn(tam_data, tam_path, tam_blocks,
  modes = tam_modes, boot.val = TRUE, br = 500)
```

본 튜토리얼에서 제시한 조형모형의 타당도의 평가 기준은 1) 가중치(weights)의 유의도 수준(t -값 > 1.96), 2) 측정지표들 간의 다중공선성 점검(VIF < 5.0) 그리고 3) Gefen -Straub의 판별 타당도 평가 기준이다. 첫째 측정지표들의 가중치의 유의성 수준 평가를 위한 t -값을 생성·확인하기 위해서는 다음과 같은 명령어들을 수행한다.

```
# 관측변수들 weights의 t-통계 확인
> tam_pls$boot$weights$t_value = with(tam_pls$
  boot $weights, Original/Std.Error)
> tam_pls$boot$weights
```

<그림 11>의 오른쪽에 보이는 바와 같이 지각된 사용용이성(PEOU)개념의 측정지표들 중 단지 peou2 가중치의 t -값만이 기준치인 1.96을 초과하고 있음을 보여 주고 있다. 즉, 조형지표들로서의 타당도에 문제점이 있음을 나타내고 있다.

둘째, 측정지표들 간의 다중공선성 점검을 위하여 측정지표들의 분산팽창계수(VIF)를 산출

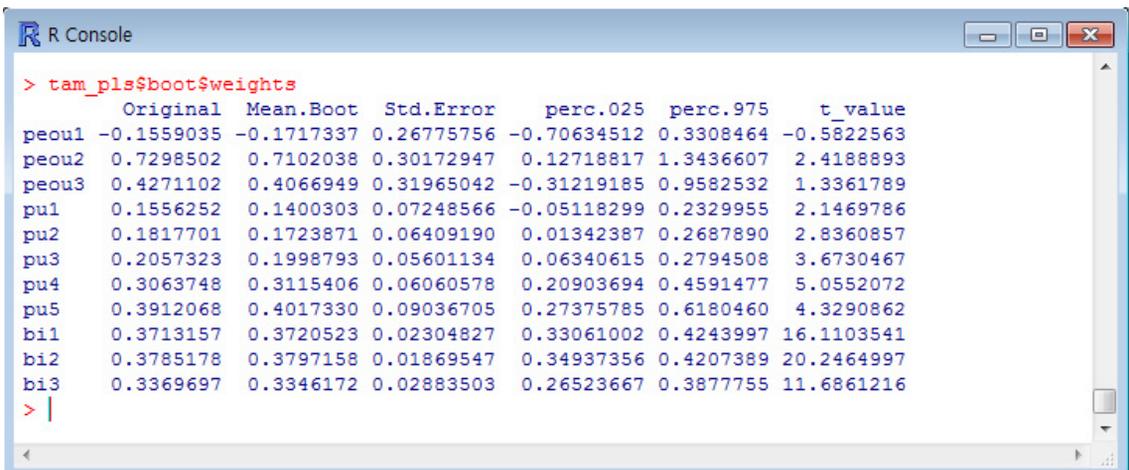
하기 방식은 다음과 같다(Hair *et al.*, 2014).

1. 첫 번째 조형지표를 종속변수(peou1)로 하고 동일 개념을 측정하기 위한 나머지 측정지표들(peou2, peou3)은 독립변수로 한 회귀분석을 시행하여, 독립변수들과 연관된 종속변수(peou1)의 분산을 확인한다(R^2_{peou1})
2. $1-R^2_{peou1}$ 을 이용하여 첫 번째 조형지표인 peou1의 공차를 계산한다.
3. 공차를 근간으로 분산팽창계수(VIF)를 산출한다(분산팽창계수(VIF) = $1/\text{공차}$). 상기의 1단계를 수행하기 위한 R에서의 회귀분석은 다음과 같다.

```
> reg_peou1 <- lm(peou1~peou2+peou3, data=
  tam_data)
> summary(reg_peou1)
```

<그림 12>에서 R^2_{peou1} 의 값이 0.4199임을 확인할 수 있다. 확인된 R^2_{peou1} 을 근간으로 아래와 같은 분산팽창계수(VIF)를 산출하는 수식을 직접 입력하여 계산한다.

```
> vif_peou1=1/(1-0.4199)
> vif_peou1
```



	Original	Mean.Boot	Std.Error	perc.025	perc.975	t_value
peou1	-0.1559035	-0.1717337	0.26775756	-0.70634512	0.3308464	-0.5822563
peou2	0.7298502	0.7102038	0.30172947	0.12718817	1.3436607	2.4188893
peou3	0.4271102	0.4066949	0.31965042	-0.31219185	0.9582532	1.3361789
pu1	0.1556252	0.1400303	0.07248566	-0.05118299	0.2329955	2.1469786
pu2	0.1817701	0.1723871	0.06409190	0.01342387	0.2687890	2.8360857
pu3	0.2057323	0.1998793	0.05601134	0.06340615	0.2794508	3.6730467
pu4	0.3063748	0.3115406	0.06060578	0.20903694	0.4591477	5.0552072
pu5	0.3912068	0.4017330	0.09036705	0.27375785	0.6180460	4.3290862
bi1	0.3713157	0.3720523	0.02304827	0.33061002	0.4243997	16.1103541
bi2	0.3785178	0.3797158	0.01869547	0.34937356	0.4207389	20.2464997
bi3	0.3369697	0.3346172	0.02883503	0.26523667	0.3877755	11.6861216

<그림 11> 가중치들의 t 값

```

R Console
> reg_peou1 <- lm(peou1~peou2+peou3, data=tam_data)
> summary(reg_peou1)

Call:
lm(formula = peou1 ~ peou2 + peou3, data = tam_data)

Residuals:
    Min       1Q   Median       3Q      Max
-2.8566 -0.4733  0.2743  0.6624  2.1624

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.5472     0.4037   3.832 0.000257 ***
peou2        0.3833     0.1301   2.946 0.004251 **
peou3        0.2524     0.1212   2.082 0.040656 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.011 on 77 degrees of freedom
Multiple R-squared:  0.4199,    Adjusted R-squared:  0.4048
F-statistic: 27.87 on 2 and 77 DF,  p-value: 7.841e-10

> |
    
```

<그림 12> R²의 확인을 위한 회귀분석 결과

```

R Console
> vif_peou1=1/(1-0.4199)
> vif_peou1
[1] 1.723841
> |
    
```

<그림 13> 측정지표 peou1의 분산평창계수

<그림 13>에서 보이듯이 peou1의 분산평창계수는 1.72로 나타나고 있다. 다른 측정지표들도 상기와 같은 방법으로 분산평창계수를 산출하면 peou2는 2.75, peou3는 2.61로 확인되며, 지표들 간의 다중공선성은 없는 것으로 평가된다.

마지막으로 조형모형의 판별 타당도 평가 방법은 이전 장의 두 번째의 판별 타당도 평가방법을 참조하기 바란다.

4.3 내부모형의 평가

예시 연구모형의 PLS 실행을 통한 내부모형에 대한 평가는 경로계수, 결정계수 R² 및 GoF에 대한 평가 순으로 진행된다.

4.3.1 경로분석

내부모형의 평가에서 경로계수의 평가는 잠

```

R Console
> tam_pls$boot$paths
      Original Mean.Boot Std.Error perc.025 perc.975 t_value
PEOU -> PU 0.3454880 0.3654991 0.10250931 0.1655193 0.5501143 3.370308
PEOU -> BI 0.3157133 0.3185316 0.09926121 0.1193374 0.4997027 3.180632
PU -> BI   0.4694681 0.4788678 0.09298639 0.3203402 0.6536953 5.048783
> |

```

〈그림 14〉 경로계수와 t- 값

```

R Console
> tam_pls$boot$rsq
      Original Mean.Boot Std.Error perc.025 perc.975
PU 0.1193619 0.1440768 0.07051227 0.02991414 0.3026266
BI 0.4224898 0.4581750 0.08526889 0.28151010 0.6039493
> |

```

〈그림 15〉 결정계수(R^2)값

재변수들간의 인과관계에 대한 평가로서 경로 계수의 통계적 유의성 수준($p < 0.05$)으로 평가 된다. 앞서 측정지표들의 집중타당도 평가에서도 언급하였듯이 부트스트랩에 의한 추정치에 대한 유의성에 대한 검증은 신뢰구간에 근간한 검증으로 수행되는 것이 논리적이다. 그러나 기존의 연구들에서 t-값을 근간으로 기준을 제시하고 있으므로 이전과 같이 t-값을 생성·추가할 필요가 있다. 경로계수와 유의성수준(p-값)를 평가하기 위한 t-값을 산출하기 위한 명령은 다음과 같으며, 그 결과는 <그림 15>와 같다.

```

# t-value 생성(추정치/오차)
> tam_pls$boot$paths$t_value = with(tam_pls$boot
  $paths, Original/Std.Error)

# 경로계수와 t-value 확인
> tam_pls$boot$paths

```

예시 연구모형의 경우 <그림 14>에서 보는 바와 같이 잠재변수들의 모든 경로계수의 t-값이

3.1 이상으로 나타나 통계적으로 유의($p < 0.01$)한 것으로 해석된다. 비록 t-값을 통한 경로계수 검증이 사회과학연구들에서 일반적으로 이루어지고 있는 실정이나 95%의 신뢰구간의 양쪽기준이 되는 perc. 025와 perc. 975의 값들을 제공하는 것이 바람직하다. 이러한 정보들은 추정치가 분포상의 어느 정도에 위치에 있는지와 같은 보다 구체적인 정보를 제공한다는 의미를 가지고 있다(Hair *et al.*, 2012b, Henseler *et al.*, 2009)

4.3.2 결정계수(R^2)

내부모형의 평가에서 R^2 에 대한 평가는 내생 잠재변수에 대한 외생 잠재변수들의 설명력을 보여주는 평가로 R^2 값이 클수록 외생 잠재변수의 설명력이 높아짐을 의미한다. 회귀분석과 같이 일반적으로 외생 잠재변수의 수가 증가할수록 R^2 의 값이 높아지는 현상을 보이며 이에 대한 명확한 기준은 없고 연구영역에 따라 다르게 제시되고 있는 실정이다. R에서 종속변수들의 R^2 값을 보기 위한 명령은 다음과 같으며, 그 산출결과는 <그림 15>와 같다.

```
> tam_pls$boot$rsq
```

예시 연구모형에서는 <그림 15>에서 보는 바와 같이 내생 잠재변수인 지각된 유용성(PU) 변수와 행위의도(BI) 변수의 R^2 값이 각각 0.11 및 0.42으로 나타나고 있다.

4.3.3 GoF

“plsmp”에서는 모형적합도 판정을 위해 GoF (Goodness-of-Fit) 값을 제공하고 있다. GoF을 보기 위한 명령은 다음과 같으나 적합도 판정을 위한 Gof 기준값이 아직 명확히 제시 되지 못하고 있기 때문에 PLS 분석결과 산출시에 GoF 결과값은 참고 수준에서만 제공될 필요가 있다.

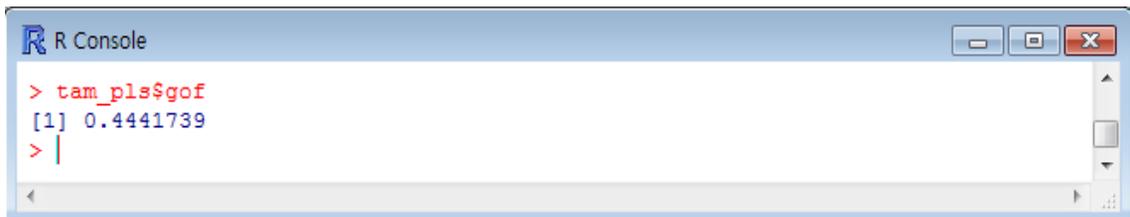
```
> tam_pls$gof
```

V. 결 론

본 튜토리얼은 최근 경영학을 비롯한 제반 사회과학연구들에서 PLS 구조방정식모형에 기반한 통계분석 연구들이 급증하고 새로운 통계 패키지 환경인 R이 폭넓게 적용되고 있는 현실을 감안하여 R 패키지인 “plsmp”을 이용하여 구조방정식모형을 분석하는 구체적인 절차와 방법을 제시하고자 하였다. 이를 위해 본 튜토리얼에서는 기존 연구들에서 제시된 PLS 구조방정식모형을 분석하는데 적용되어야 하는 평가항목 및 평가기준들을 통합적으로 정

리하여 제시한 후 경영정보시스템 분야 연구에서 가장 잘 알려진 이론적 모형들 중 하나인 기술수용모형(TAM)을 예시 연구모형으로 하고 본 연구모형에 포함된 변수들에 대한 실제 수집된 데이터를 대상으로 해 R을 이용하여 실질적으로 연구모형을 PLS 구조방정식모형으로 분석하는 과정을 R 명령어와 산출물을 중심으로 보여주었다. 또한 본 튜토리얼에서 제시한 절차와 방법대로 연구자들이 실제로 수행해 볼 수 있도록 <부록 1>에 분석자료(TAM.csv)를 첨부하였다.

본 튜토리얼이 PLS 분석을 처음 접하는 연구자들에게는 구조방정식 연구모형 분석을 보다 체계적이고 적절하게 수행하기 위한 유용한 사용지침서가 될 것으로 예상하며, 아울러 PLS에 이미 익숙한 연구자들에게는 R이라는 새로운 통합 통계소프트웨어 운영환경에서 보다 창의적이고 심도 있는 연구분석을 할 수 있는 계기를 제공할 수 있을 것으로 기대한다. 비록 R을 이용한 통계분석이 창의적이고 심도 있는 연구분석을 가능케 할 수 있으나, 이를 효과적으로 수행하기 위해서는 컴퓨터 프로그래밍 언어에 대한 기본 지식을 습득하는 것이 이 바람직하다. 연구자들은 이러한 점을 명심하기 바라며, 아울러, 본 튜토리얼이 경영정보시스템 연구자들의 R에 대한 적극적인 활용과 컴퓨터 프로그래밍 언어에 대한 기본 지식을 습득하는 기회요인으로서의 계기가 되기를 바란다.



<그림 16> GoF 값

참고 문헌

- 김광섭, 이기원, “R 이용 오픈데이터 시각화 웹 응용”, 한국지리정보학회지, 제17권, 제2호, 2014, pp. 72-81.
- Addinsoft, XLSTAT-Statistics Package for Excel, 2011, URL <http://www.xlstat.com/>.
- Bagozzi, R. P. and Y. Yi, “On the Evaluation of Structural Equation Models”, *Journal of the Academy of Marketing Science*, Vol.16, No.1, 1988, pp. 74-94.
- Chin, W. W., “Issues and opinion on structural equation modeling”, *MIS Quarterly*, Vol.22, No.1, 1998a, pp. 7-16.
- Chin, W. W., The partial least squares approach to structural equation modeling. In: Marcoulides, G. A.(Ed.), *Modern Methods for Business Research*. Lawrence Erlbaum Associates, Mahwah, NJ, 1998b, pp. 295-336.
- Chin, W. W., PLS Graph-Version 3.0. Soft Modeling Inc. 2003, URL <http://www.plsgraph.com/>.
- Davis, F. D., “Perceived usefulness, perceived ease of use, and user acceptance of information technology”, *MIS Quarterly*, Vol.13, No.3, 1989, pp. 319-340.
- Diamantopoulos, A. and J. A. Siguaw, “Formative vs. reflective indicators in organizational measure development: A comparison and empirical illustration”, *British Journal of Management*, Vol.17, No.4, 2006, pp. 263-282.
- Fornell, C. and D. F. Larcker, “Evaluating Structural Equation Models with Unobservable Variables and Measurement Error”, *Journal of Marketing Research*, Vol.18, No.1, 1981, pp. 39-50.
- Fox, J. and R. Anderson, Using the R Statistical Computing Environment to Teach Social Statistics Courses. 2005, <http://www.unt.edu/rss/Teachingwith-R.pdf>.
- Fu, J.-R. VisualPLS-Partial Least Square(PLS) Regression-An Enhanced GUI for Lvpls (PLS 1.8 PC) Version 1.04. National Kaohsiung University of Applied Sciences, Taiwan, ROC, 2006.
- Gefen, D. and D. W. Straub, “A Practical Guide to Factorial Validity Using PLS-Graph: Tutorial and Annotated Example”, *Communications of the AIS*, Vol.16, No.1, 2005, pp. 91-109.
- Hair, J. F., G. T. M. Hult, C. M. Ringle, and M. Sarstedt(2014), *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)*. Thousand Oaks: Sage.
- Hair, J. F., Jr, M. Sarstedt, T. M. Pieper, and C. M. Ringle, “The use of partial least squares structural equation modeling in strategic management research: A review of past practices and recommendations for future applications”, *Long Range Planning*, Vol.45, No.5-6, 2012a, pp. 320-340.
- Hair, J. F., Jr., M. Sarstedt, C. M. Ringle, and J. A. Mena, “An assessment of the use of partial least squares structural equation modeling in marketing research”, *Journal of the Academy of Marketing Science*, Vol.40, No.3, 2012b, pp. 414-433.
- Henseler, J., C. M. Ringle, and R. R. Sinkovics, “The use of partial least squares path modeling in international marketing”, *Advances in International Marketing*, Vol.20, No.1, 2009, pp. 277-320.
- Henseler, J. and M. Sarstedt, “Goodness-of-fit indices for partial least squares path modeling”, *Computational Statistics*, Vol.28, No.2, 2013, pp. 565-580.
- Jarvis, C. B., S. B. Mackenzie, P. M. Podsakoff, D. G. Mick, and W. O. Bearden, “A critical review of construct indicators and measurement model misspecification in marketing and consumer research”, *Journal of Consumer Research*, Vol.30, No.2, 2003, pp. 199-218.

- Lee, L., S. Petter, D. Fayard, and S. Robinson, "On the Use of Partial Least Squares Path Modeling in Accounting Research", *International Journal of Accounting Information Systems*, Vol.12, No. 4, 2011, pp. 305-328.
- Li, Y., PLS-GUI-Graphic User Interface for Partial Least Squares (PLS-PC 1.8)-Version 2.0.1 beta. University of South Carolina, Columbia, SC, 2005.
- Lohmöller, J. B., PLS-PC: Latent Variables Path Analysis with Partial Least Squares-Version 1.8 for PCs under MS-Dos, 1987.
- MacKenzie, S. B., P. M. Podsakoff and C. B. Jarvis, "The problem of measurement model misspecification in behavioral and organizational research and some recommended solutions", *Journal of Applied Psychology*, Vol.90, No.4, 2005, pp. 710-730.
- Monecke, A. and F. Leisch, "semPLS: Structural Equation Modeling Using Partial Least Squares", *Journal of Statistical Software*, Vol.48, No.3, 2012, pp. 1-32.
- Nunnally, J. C. and I. H. Bernstein, *Psychometric Theory*, 3rd ed. McGraw-Hill, New York. 1994.
- Penga, D. X. and F. Lai, "Using partial least squares in operations management research: A practical guideline and summary of past research", *Journal of Operations Management*, Vol.30, 2012, pp. 467-480.
- Petter, S., D. W. Straub, and A. Rai, "Specifying formative constructs in information systems research", *MIS Quarterly*, Vol.31, No.4, 2007, pp. 623-656.
- Ringle, C. M., M. Sarstedt, and D. W. Straub, "Editor's Comments: A Critical Look at the Use of PLS-SEM in MIS Quarterly", *MIS Quarterly*, Vol.36, No.1, 2012, pp. iii-xiv.
- Ringle, C. M., S. Wende, and A. Will, *SmartPLS-Version 2.0*. Universität Hamburg, Hamburg. 2005.
- Rodgers, W. and A. Guiral, "Potential Model Misspecification Bias: Formative Indicators Enhancing Theory for Accounting Researchers", *The International Journal of Accounting*, Vol.46, No.1, 2011, pp. 25-50.
- Sanchez, G., *PLS Path Modeling with R* Trowchez Editions. Berkeley, 2013. [http://www.gastonsanchez.com/PLS Path Modeling with R.pdf](http://www.gastonsanchez.com/PLS_Path_Modeling_with_R.pdf).
- Tenenhaus, M., V. V. Esposito, Y. M. Chatelin, and C. Lauro, "PLS Path Modeling", *Computational Statistics and Data Analysis*, Vol.48, 2005, pp. 159-205.
- Urbach, N. and F. Ahlemann, "Structural Equation Modeling in Information Systems Research Using Partial Least Squares", *Journal of Information Technology Theory and Application*, Vol.11, No. 2, 2010, pp. 5-40.
- Wikipedia, R(programming language), [http://en.wikipedia.org/wiki/R_\(programming_language\)](http://en.wikipedia.org/wiki/R_(programming_language)), 2014.
- Wold, H., Soft modeling: the basic design and some extensions. In: Jöreskog, K. G., Wold, H. (Eds.), *Systems Under Indirect Observations: Causality, Structure, Prediction*, Part 2. North-Holland, Amsterdam, The Netherlands, 1982, pp. 1-54.
- Wong, K. K., "Partial Least Squares Structural Equation Modeling (PLS-SEM) Techniques Using SmartPLS", *Marketing Bulletin*, Vol.24, Technical Note 1, 2013, pp. 1-32.
- Yoon, C. and S. Kim, "Convenience and TAM in a ubiquitous computing environment: The case of wireless LAN", *Electronic Commerce Research and Applications*, Vol.6, No.1, 2007, pp. 102-112.

〈부록 1〉 TAM.csv

peou1	peou2	peou3	pu1	pu2	pu3	pu4	pu5	bi1	bi2	bi3
3	3	3	5	6	5	5	3	2	3	6
3	3	4	6	5	5	5	4	4	4	5
4	5	4	7	4	6	7	7	5	5	5
5	5	4	5	6	6	6	6	4	4	3
5	5	6	5	6	5	5	6	5	5	5
4	5	4	6	6	6	6	6	5	5	4
5	3	5	4	6	6	6	6	6	6	5
4	6	6	5	5	5	4	4	4	4	4
4	4	3	4	6	5	6	4	5	6	6
2	1	1	5	5	4	6	5	4	5	4
3	3	3	3	4	6	4	5	3	5	4
5	4	4	4	3	3	4	3	4	4	3
4	6	6	6	6	6	6	6	5	5	5
6	6	6	6	6	6	5	5	5	6	4
5	3	2	4	4	5	5	5	4	4	4
4	5	5	6	5	6	6	6	5	5	4
4	4	3	2	2	2	2	5	7	6	5
5	4	5	5	5	3	4	4	5	5	5
1	1	1	2	3	3	2	2	1	1	1
5	5	5	6	5	5	5	5	4	4	4
4	3	4	5	5	4	5	4	4	3	3
5	5	6	6	6	6	7	6	7	7	7
5	5	5	5	3	3	3	5	4	3	3
4	5	4	6	4	5	6	6	5	5	5
3	3	5	5	5	5	6	7	7	7	7
7	6	7	7	7	7	7	7	7	6	5
7	7	7	7	7	7	7	7	7	7	7
4	5	6	6	4	5	6	6	5	6	5
5	4	4	5	5	4	5	5	5	5	5
5	4	4	5	4	3	4	4	4	4	4
4	6	7	7	5	7	7	7	7	7	7
4	5	4	5	6	6	6	5	6	5	7
5	6	6	6	6	7	7	7	7	7	7
6	5	7	6	7	7	7	7	7	7	7
4	4	3	6	5	5	5	4	4	4	4
3	5	6	4	3	3	5	6	6	6	4
4	4	4	5	4	5	5	5	5	6	6
6	5	5	5	4	4	3	3	6	6	5
2	3	4	3	4	5	3	2	5	4	3

2	2	4	2	3	3	2	2	3	3	4
4	3	3	6	5	5	3	3	5	4	5
4	4	4	6	6	6	6	6	6	6	6
5	4	5	7	6	6	4	4	4	4	3
5	4	5	7	6	6	4	4	4	4	3
6	6	6	6	4	5	4	4	7	7	7
4	2	4	6	3	4	4	5	3	2	2
6	6	7	1	1	1	4	6	7	7	7
3	4	6	5	6	6	6	6	6	6	6
3	4	3	5	5	4	7	4	4	4	4
6	6	7	4	2	2	6	4	1	2	2
5	6	6	5	6	6	6	5	5	5	6
5	3	6	3	3	3	3	3	5	5	3
6	4	6	6	6	6	7	7	7	7	7
4	4	4	3	3	4	4	3	4	4	4
4	3	4	6	6	6	6	5	5	5	5
6	4	3	6	6	6	2	2	2	2	4
6	6	7	4	5	5	5	6	6	6	4
5	5	5	6	6	6	6	5	6	7	7
1	2	3	7	6	6	6	3	6	7	2
4	3	3	7	7	7	7	7	6	6	6
4	6	5	5	4	4	4	3	5	5	6
6	6	7	5	5	6	5	4	7	5	5
5	4	4	3	3	4	5	5	5	4	5
2	6	4	6	5	4	4	5	7	7	4
5	5	5	4	4	6	4	4	3	3	4
6	6	6	4	4	4	3	4	6	6	6
5	4	4	6	6	5	5	6	7	7	6
3	5	6	4	5	5	6	6	7	7	6
3	3	3	7	7	7	7	4	5	5	6
7	7	7	5	5	6	6	7	7	7	7
4	7	7	4	4	4	7	5	7	7	7
7	7	7	6	4	4	5	7	7	7	3
5	5	3	6	6	6	4	3	7	7	7
4	4	5	3	2	4	4	4	1	2	4
5	4	5	5	5	6	6	5	5	4	4
4	5	4	4	5	6	7	6	6	6	5
4	2	2	3	3	3	2	2	3	3	2
5	5	5	7	7	6	6	6	6	6	6
3	3	4	7	6	4	5	4	4	4	4
6	5	5	6	5	5	3	4	4	3	4

Information Systems Review

Volume 16 Number 3

December 2014

A Tutorial on PLS Structural Equating Modeling using R: (Centering on) Exemplified Research Model and Data

Cheolho Yoon* · Sanghoon Kim**

Abstract

This tutorial presents an approach to perform the PLS structural equation modeling using the R. For this purpose, the practical guide defines the criteria for the PLS structural equation modeling by reviewing previous studies, and shows how to analyze the research model with an example using the “plspm” which is the R package for the performing PLS path analysis against the criteria. This practical guide will be useful for the study of the PLS model analysis for new researchers and will provide the knowledge base for in-depth analysis through the new PLS structural equation modeling technique using R which is the integrated statistical software operating environment for the researchers familiar with the PLS structural equation modeling.

Keywords: *R Programming, Structural Equating Modeling(SEM), Partial Least Squares(PLS), Technology Acceptance Model(TAM), Plspm, SmartPLS, Path Modeling*

* Corresponding Author, Associate Professor, Department of Business Administration, Mokpo National University

** Professor, School of Business, Kwangwoon University

◎ 저 자 소 개 ◎



윤 철 호 (carlyoon@empal.com)

현재 국립목포대학교 경영학과 부교수로 재직 중이며, 광운대학교 전자계산학과에서 학사, 경영정보학과에서 박사를 취득하였다. Information and Management, Journal of Computer Information Systems, Journal of Business Ethics, Electronic Commerce Research and Applications, Computers in Human Behavior, Behaviour and Information Technology, and Journal of Organizational Computing and Electronic Commerce 등 십여 편의 국제학술지와 경영학연구, 경영정보학연구, 한국경영과학회지, 경영과학, Information Systems Reviews 등의 국내학술지에 논문을 게재한 바 있으며, 현재 Information Technology and Management의 편집위원과 International Journal of Information Management 등 십여 개의 국제저널에서 리뷰어로 활동하고 있다.



김 상 훈 (shkim@kw.ac.kr)

현재 광운대학교 경영학부 교수로 재직 중이며, 서울대학교 경제학과를 졸업하고 한국과학기술원(KAIST) 경영과학과에서 석사 및 박사학위를 취득하였다. 주요 관심분야는 정보화 전략 수립 및 추진, 정보시스템실행을 위한 변화관리, IT 활용에 의한 경영혁신, 정보시스템평가, ERP(Enterprise Resource Planning) 시스템 실행, S/W개발프로젝트관리 등이다.

논문접수일 : 2014년 08월 25일

게재확정일 : 2014년 11월 25일

1차 수정일 : 2014년 11월 19일