

心理調査におけるパス解析のためのRスクリプト

R script for Path Analysis in psychological survey

緒賀 郷志 (岐阜大学教育学部) Satoshi OGA

はじめに

心理学研究においては様々な研究方法が用いられる。それは大別すると実験法と観察法となる。観察法の中でも調査研究は卒業研究や修士論文研究で多く用いられている。その調査研究のデータを分析するための統計的手法も多く開発されているが、本稿ではパス解析をとりあげて、その分析手順を統計環境ソフトであるR (R core team, 2014) でもって解説する。従来、心理学研究におけるデータ分析ソフトとして最も使われてきたのはSPSSであろう。そのようななか、Rはオープンソースソフトであり、学部学生ならびに大学院生が自分の所持するパソコンにインストールし、自由に活用できるため、これからはますますRの利用頻度は高まっていくと考えられる。しかしながら、Rは基本的にコマンド入力による操作のため、初心者にはハードルが高い面がある。そこで本稿では、調査データの基本的な分析としてのパス解析を実行するスクリプトを掲載するとともに、コマンドの解説を加える事で、初心者取り組みやすくなることを期待している。

Rスクリプトと解説

まず前提として、data.csvのファイルに以下の9つの変数が入っていることとする。ID, sex, age, A, B1, B2, C, Dである。sexの変数は、男性はM, 女性はFでコーディングされているとする。またそれぞれの変数に欠損値がないという前提で分析をすすめる。

CSVファイルから、R環境内のデータフレームへの読み込み方法については次のスクリプトを使う。

```
(wd<- getwd( ))
x <- read.csv(paste0(wd,"/data.csv"))
```

getwd関数は、現在の作業ディレクトリ (フォルダ) の所在地 (パス) を得る。それをwdという名称のオブジェクト (wdという名称でなくてもよい) に代入する。()でくくってあるのは、その実行結果が即時に得られるようにするためである。次のread.csv関数は、csvファイルに格納されているデータを読み込む。その際に引数として、所在地を含めた形でファイル名を指定する。そのスクリプトが、paste0(wd,"/data.csv")である。paste0は、引数に有る文字列を結合するが、その際に結合箇所に空白をいれない関数である。上記のスクリプトを走らせることで、オブジェクト名xに、分析するデータが読み込まれる。この分析するデータが格納されているxはデータフレームと呼ばれる。

次は基本統計量を確認するスクリプトである。

```
dim(x)                # 行と列の数を調べる (人数と変数の数)
anyNA(x)              # 欠損値のあるか確認 FALSE とではず TRUEならどこかに欠損値有り
(nsex <- with(data = x, table(sex))) # 男女の人数を知る
chisq.test(nsex)      # 男女の人数に差がないか  $\chi^2$ 検定
library(psych)        # psych パッケージの呼び出し
with(x, describe(age)) # 年齢についての基本統計量を確認する
with(x, table(sex, age)) # 性別と年齢のクロス表をつくってみる
with(x, boxplot(age ~ sex)) # 男女別の年齢の箱ひげ図を描く
with(x, t.test(age ~ sex)) # 男女で年齢差がないか Welch の検定
with(x, describeBy(age, group=sex))
describeBy(x[, 3:9], group=x$sex)      # 男女別の基本統計量
```

各コマンド (スクリプト) の後ろに、#記号があり、コメントの文が書かれている。R では、#以降の文は改行があるまでは、コメントの文章として実行されない。

dim 関数は、dimension (次元) の略であり、行列の数を返してくる。この場合、最初に返ってくるのが行数 (人数) であり、2 番目の数が列数 (変数の数) となる。anyNA 関数は、欠損値 NA があるかを調べる関数である。欠損値がなければ FALSE となる。ここで TRUE ならば、欠損値の処理をどうするかを検討した上で分析を進めなければならない。次に、男性と女性のそれぞれの人数について調べる。table 関数で集計ができる。その時に、table(x\$sex) とすることで、x のデータフレームにある中の変数 sex を直接に指定することもできるが、今回、with 関数を用いている。with 関数で引数、data = x と指定することで、そのあとの変数名は、x\$何々 という形をとらないで済むので便利である。with 関数の最初の引数にデータフレームのオブジェクト名を指定することで、data= は省略できる。男女の人数を知ると同時に、その結果をnsexというオブジェクトに代入をしているので、男女の人数に差がないかの χ^2 検定を chisq.test 関数で実行できる。出力には χ^2 の値が、X-squared = で、自由度は df = で、有意確率は p-value = で示される。

library 関数は R の基本となるコマンド以外が収納されている追加パッケージを読み込むための関数である。この場合、psych パッケージを読み込むことで、基本統計量を算出する describe 関数が利用可能となる。年齢に関する基本統計量を算出するのは、with(x, describe(age)) であるが、x のオブジェクト内にあるすべての変数の基本統計量を算出するのに、with(x, describe(x)) とすることができる。describe 関数では、人数である n、平均値である mean、標準偏差である sd、中央値 median、最小値 min、最大値 max などが算出される。ただし x を引数とした時、sex の変数については、数値でない factor 型なため、計算不可能の警告メッセージがでるので注意されたい。さて、table 関数はクロス表も作成できる。x 内にある sex と age のクロス表は、with(x, table(sex, age)) で作成できるし、同様に箱ひげ図は boxplot 関数で描画できるし、t 検定は t.test 関数で実行できる。その際の引数の形式は、~ チルダ記号を用い、チルダの前の変数を、チルダの後の変数で説明分割する。なお t.test の結果ではそれぞれのグループの平均は算出表示されるが、標準偏差が表示されないのので、別途に計算する必要がある。それに用いるのが psych パッケージ内の describeBy 関数である。group = sex という引数を追加することで、男女別の基本統計量が算出される。sex の変数に関する警告メッセージを見たくなければ、x のデータフレームの 3 列目から 8 列目を指定して分析をす

ることが可能である。x[,3:8] がその指定方法である。[]内の、カンマの前は行を指定する部分であるが、今回は何も指定しない。カンマのあとのコロンの結んだ3:8が3列目から8列目を指定している。さてパス解析の前に、変数間の散布図を描き、相関係数表を算出することとする。

```
pairs.panels(x[, c("A","B1","B2","C","D")]) # 全体の散布図, x[, 4:8]としても良い。
pairs.panels(x[x$sex=="F", c("A","B1","B2","C","D")]) # 女性の散布図
pairs.panels(x[x$sex=="M", c("A","B1","B2","C","D")]) # 男性の散布図
```

男女込みの相関係数

```
cp <- corr.test(x[, c("A","B1","B2","C","D")]) # 引数はx[, 4:8]としても良い。
print(cp, digits=3, short=FALSE)
# 相関係数および無相関検定と、95%信頼区間の表示
```

女性だけの相関係数

```
cp <- corr.test(x[x$sex=="F", c("A","B1","B2","C","D")])
print(cp, digits=3, short=FALSE) # 相関係数および無相関検定と、95%信頼区間の表示
```

男性だけの相関係数

```
cp <- corr.test(x[x$sex=="M", c("A","B1","B2","C","D")])
print(cp, digits=3, short=FALSE) # 相関係数および無相関検定と、95%信頼区間の表示
```

pairs.panels 関数も、psych パッケージ内の関数である。x のデータフレーム内の A, B1, B2, C, D の6つの変数間の関連性をみるため、x[, c("A","B1","B2","C","D")] を引数としているが、これらの変数は x のデータフレームの4列目から8列目なので、x[, 4:8] としても良い。つぎに、女性だけの散布図、男性だけの散布図を出して見る際に、x のデータフレームの行を指定することで行うことができる。すなわち、女性のみならば、行指定に、x\$sex=="F" を設定し、男性のみならば、行指定に x\$sex=="M" を設定するのである。ようするに、x[x\$sex=="F", 4:8] というのは sex 変数が女性 F である行のみ、かつ、4列目から8列目の変数を取り出しているということである。pairs.panels 関数でも相関係数が算出・表示されるが、無相関検定は行われない

相関係数表ならびに無相関検定の結果は、corr.test 関数で算出できる。corr.test の結果を cp というオブジェクト名に代入し、print(cp, digits=3, short=FALSE) とすることで、相関表 Correlation matrix と、分析した人数 Sample Size が表示される。さらに、無相関検定の確率の値 Probability values (Entries above the diagonal are adjusted for multiple tests.) が出るとともに、Confidence intervals based upon normal theory. という形で、相関係数の95%信頼区間が表示される。

全体と男女別の結果を見比べて、同じ傾向にあると考えれば、全体で分析を行えばよいだろうし、男女で相関の傾向が違えば、男女別にパス解析を進めたほうがいだろう。

次にパス解析である。今回は SEM を利用したパス解析を行うため、lavaan パッケージを利用して実行する。

```
##### SEM を使ったパス解析のメモ #####
# モデルの指定の方法
#
# ~ 構造方程式 (回帰) (単方向矢印 <-)
# ~~ 残差の共分散 (相関) (双方向矢印 <-> )
#
# 一重引用符で括ること
#####
```

```
library(lavaan)
sem.model.0 <- 'C ~ A + B1 + B2'
```

このsem.model0 は、C を目的変数とし、A、B1、B2 を説明変数とした重回帰分析のモデル指定である。その後、下記のように lavaan パッケージの sem 関数を用いて計算を実行し、summary 関数でその結果を表示させる。

```
fit0 <- sem(sem.model.0, data=x)
summary(fit0, standardized=TRUE, rsquare=TRUE, fit=TRUE)
inspect(fit0, "cor.all")
```

summary 関数での引数では、standardized=TRUE の引数をつけることで、標準化推定値 (Std. all) が出力される。さらに rsquare=TRUE の引数をつけることで重相関係数の平方が出力される。また、fit=TRUE の引数で適合度が出力される。inspect(fit0, "cor.all") では相関係数表が表示される。

次にD変数に向かって、C変数から矢印を伸ばした構造方程式を追加したパス解析モデルを検討する。

```
sem.model.1 <- '
C ~ A + B1 + B2
D ~ C
'
fit1 <- sem(sem.model.1, data=x)
summary(fit1, standardized=TRUE, rsquare=TRUE, fit=TRUE)
inspect(fit1, "cor.all")
modificationindices(fit1) # mi の値が高いパスを加えてみると適合度が高まる
```

モデル修正には、modificationindices 関数を用いると良い。その場合、mi の値が高いパスを加え

ると適合度が高まる。また有意でないパスを削除することによるモデル修正も行って良いだろう。

さて、次にD変数に向かって、A変数から直接に矢印を伸ばした構造方程式を追加した別なモデルを検討する。

```
sem.model.2 <- '
C ~ A + B1 + B2
D ~ A + C
'

fit2 <- sem(sem.model.2, data=x)
summary(fit2, standardized=TRUE, rsquare=TRUE, fit=TRUE)
inspect(fit2, "cor.all")
fitmeasures(fit2) # 適合度指標の算出
```

最後の fitmeasures 関数で適合度指標が計算・表示される。さらに次の anova 関数で、AIC 基準でのどれがベターなモデルなのかの検定が可能である。このようにして、それぞれの理論モデルと、それに対する様々な適合度指標を参考にしながら、より適切なパスモデルを検討すればよい。

```
anova(fit0, fit1, fit2)
```

なお、semPlot パッケージで図が描けるが、確認のための利用できる質でしかないため、図の作成のためには、libreoffice の draw ソフトを使うと良いだろう。

```
library(semPlot)
semPaths(fit2, "std", layout="tree2",
         fade=FALSE,           #色が薄くなるのをやめる
         gray=TRUE)           #モノクロ指定
```

男女別の検討を行う場合には、sem 関数に group="sex" の引数をつけるだけで実行できる。

```
fitting1 <- sem(sem.model.2, data=x, group="sex")
summary(fitting1, std=TRUE, fit=TRUE)
standardizedSolution(fitting1)
```

標準化係数のみを出力させるのに、standardizedSolution 関数というのがあるのも覚えておいてもよい。さて、同じモデルで男女のパス係数の値に差があるかを調べるために作成した関数 paratest

が下のスクリプトである。

```
##### 2群のパス係数の差の比較の関数 #####
paratest <- function(f){
  para <- parameterEstimates(f)
  n <- nrow(para)/2
  out.df <- para[1:n,1:3]
  out.df <- transform(out.df, z=NA, p=NA)
  for (i in 1:n){
    a1 <- para$est[i]
    s1 <- para$se[i]
    a2 <- para$est[i+n]
    s2 <- para$se[i+n]
    zp <- (abs(a1-a2)/sqrt(s1^2+s2^2))
    out.df[i,4] <- zp
    out.df[i,5] <- 2*pnorm(-abs(zp)) }
  print(out.df)
}
#####
```

パス係数のパラメータの値の出力は parameterEstimates(fitting1) で可能であるので、その値を用いて計算する。使い方は、下記の1行のスクリプトですむ。

```
paratest(fitting1)
```

終わりに

以上で、R を利用したパス解析の実行ができるようになるだろう。なお本稿で確認した R のバージョンは、3.1.2 であり、psych パッケージ (Revelle, 2015) のバージョンは1.5.6, そしてlavaan パッケージ (Yves, 2012) のバージョンは0.5-18である。これらは citation 関数または、sessionInfo 関数で調べることができる。

なおlavaan に関しては、出版物 (豊田, 2014; 小杉・清水, 2014) やオンライン上の情報も多くでてきているので、それらを参照すると良いだろう。

文献

小杉考司・清水裕士 (2014) M-plusとRによる構造方程式モデリング入門 北大路書房

R Core Team (2014). R: A language and environment for statistical computing. R Foundation for

Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.

Revelle, W. (2015) psych: Procedures for Personality and Psychological Research, Northwestern University, Evanston, Illinois, USA, <http://CRAN.R-project.org/package=psych> Version = 1.5.6.

豊田 秀樹 (2014) 共分散構造分析 R編—構造方程式モデリング 東京図書

Yves Rosseel (2012). lavaan: An R Package for Structural Equation Modeling. Journal of Statistical Software, 48(2), 1-36. URL <http://www.jstatsoft.org/v48/i02/>.

Yves Rosseel 2012年12月19日版 (日本語訳 荒木孝治 岸谷和広 馬場一 2013年1月24日)
lavaan: 構造方程式モデリングおよびその他のためのRパッケージバージョン 0.5-12 (ベータ版)
<http://www.ec.kansai-u.ac.jp/user/arakit/documents/lavaanPackageVer0.5-12.pdf>

Yves Rosseel 2013年7月21日版 (日本語訳 荒木孝治 2013年11月12日)
lavaanチュートリアル
<http://www.ec.kansai-u.ac.jp/user/arakit/documents/lavaanTutorial%20July%2021,%202013.pdf>