

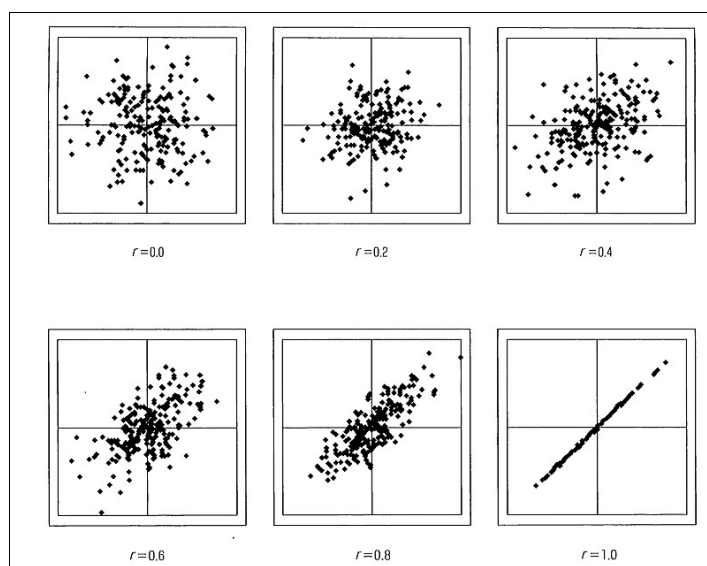
◎相関係数の概要説明

1. 相関係数の式

$$r_{xy} = \frac{s_{xy}}{s_x s_y}$$

r_{xy} （単に r とも）は相関係数（ピアソンの積率相関係数）、 s_{xy} は共分散、 s_x と s_y はそれぞれ x と y の標準偏差を表す。

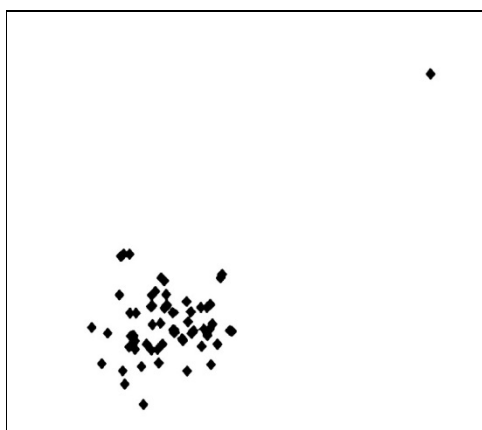
2. 相関係数の大きさ



$0 < r \leq 0.2$ または $-0.2 \leq r < 0$	ほとんど相関なし
$0.2 < r \leq 0.4$ または $-0.4 \leq r < -0.2$	弱い相関あり
$0.4 < r \leq 0.7$ または $-0.7 \leq r < -0.4$	中程度の相関あり
$0.7 < r \leq 1.0$ または $-1.0 \leq r < -0.7$	強い相関あり

3. 外れ値の影響

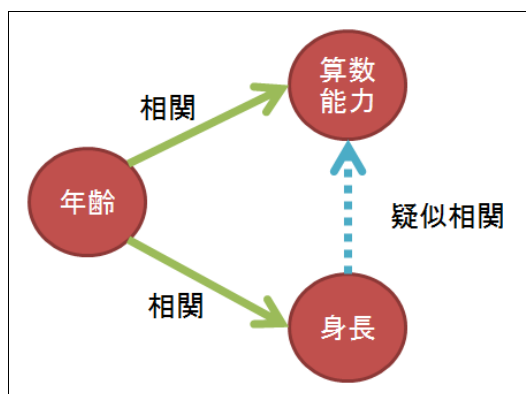
外れ値(極端に大きな、または小さな値の少数のデータ)によって、相関係数の値は大きく変化する。



(例)全体の相関が $r=0.59$ でも、右上の外れ値を取り除くと、相関はわずか $r=0.07$ になる。
→ 散布図で外れ値を確認する必要がある。

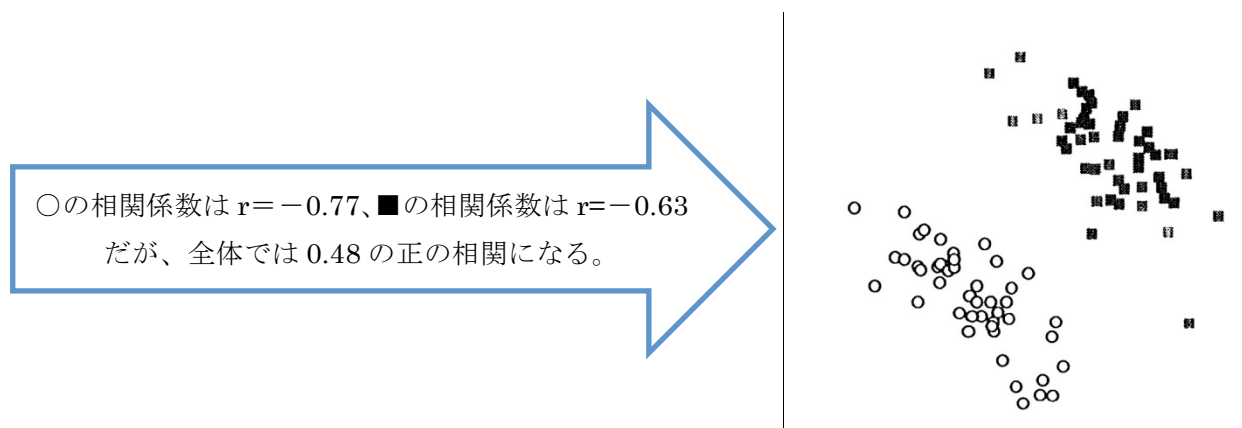
4. 疑似相関

疑似相関とは、2 つの変数 x と y の間に本当は相関がないにも関わらず、第 3 の変数 z の存在によりあらわれる見かけ上の相関のこと。



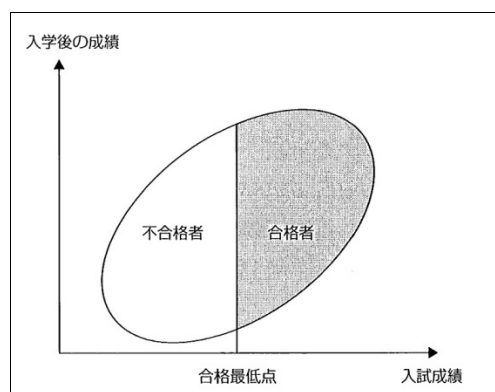
5. 層別相関

層別相関とは、集団ごとに（＝層別化した）計算した相関係数のこと。



6. 選抜効果(切断効果)

選抜効果(切断効果)とは、集団の一部が切り取られたため生じる相関の変動のこと。



7. 相関係数と因果関係

相関係数と因果関係は必ずしも一致しない。つまり、因果関係にある「 $A \rightarrow B$ 」という方向性が相関係数には存在しない。相関が強くても「 $A \rightarrow B$ 」なのか「 $B \rightarrow A$ 」なのかは分からない。

◎R コードの説明

Larson-Hall, Jenifer. (2015). "6 Correlation". *A Guide to Doing Statistics in Second Language Research Using R*. pp. 60- 82

*はじめに

<<http://cw.routledge.com/textbooks/9780805861853/spss-data-sets.asp>>

上記のサイトから、

- FlegeYeniKomshianLiu.sav
- DeKeyser2000.sav
- LarsonHallPartial.sav
- LarsonHallGJT.sav

を、それぞれダウンロードする。

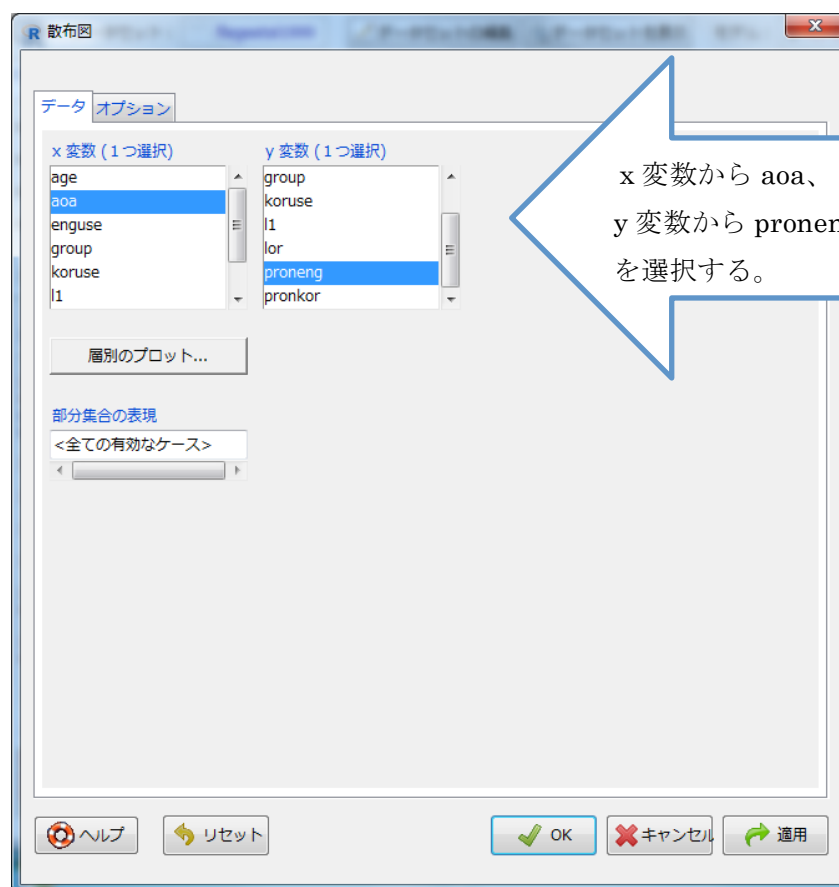
6.1 散布図の作成

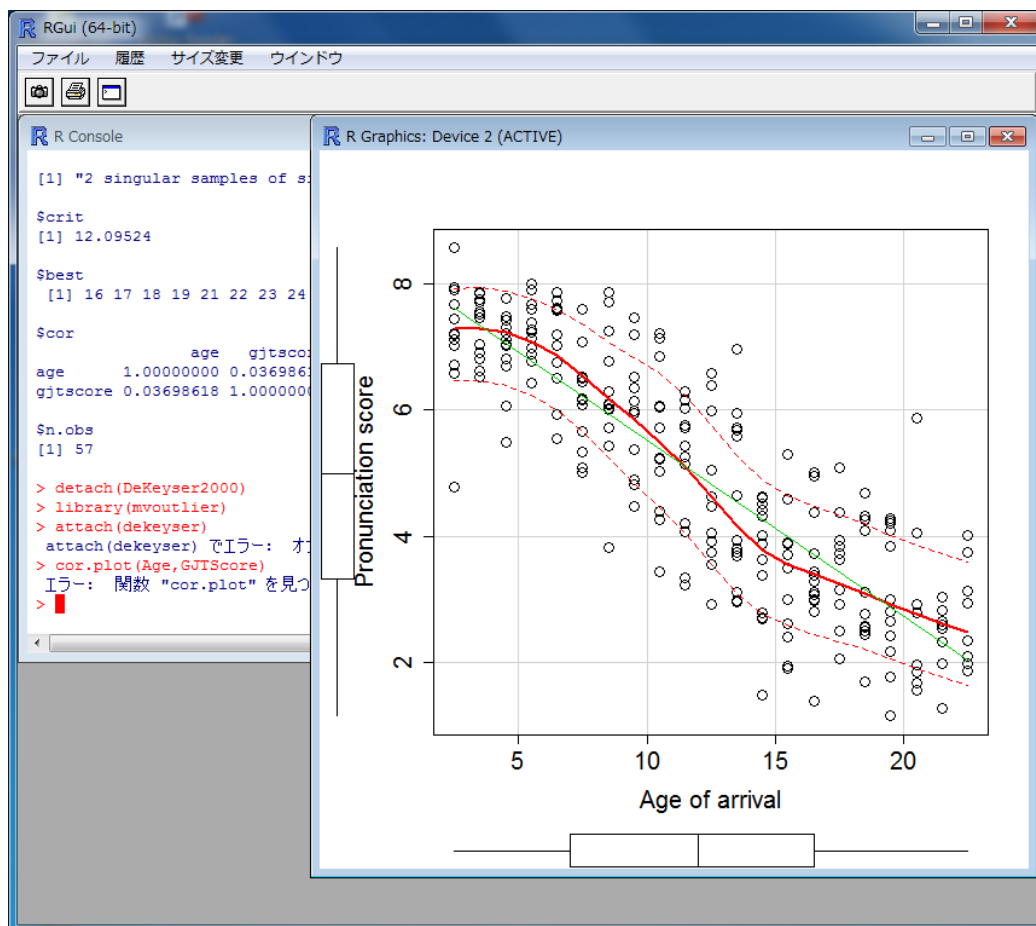
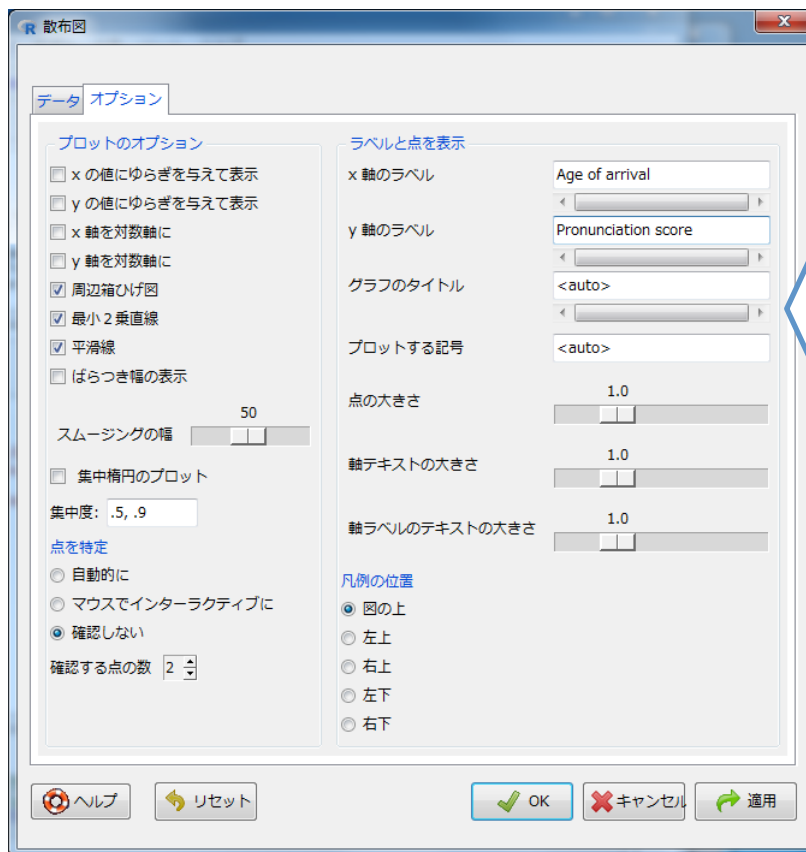
①Data: FlegeYeniKomshianLiu.sav

Name: flegeetal1999

R Cmdr: グラフ > 散布図

R Code: `library(car) #this activates the car library, which has the scatterplot command`
`scatterplot(proneng~aoa, data=flegeetal1999)`





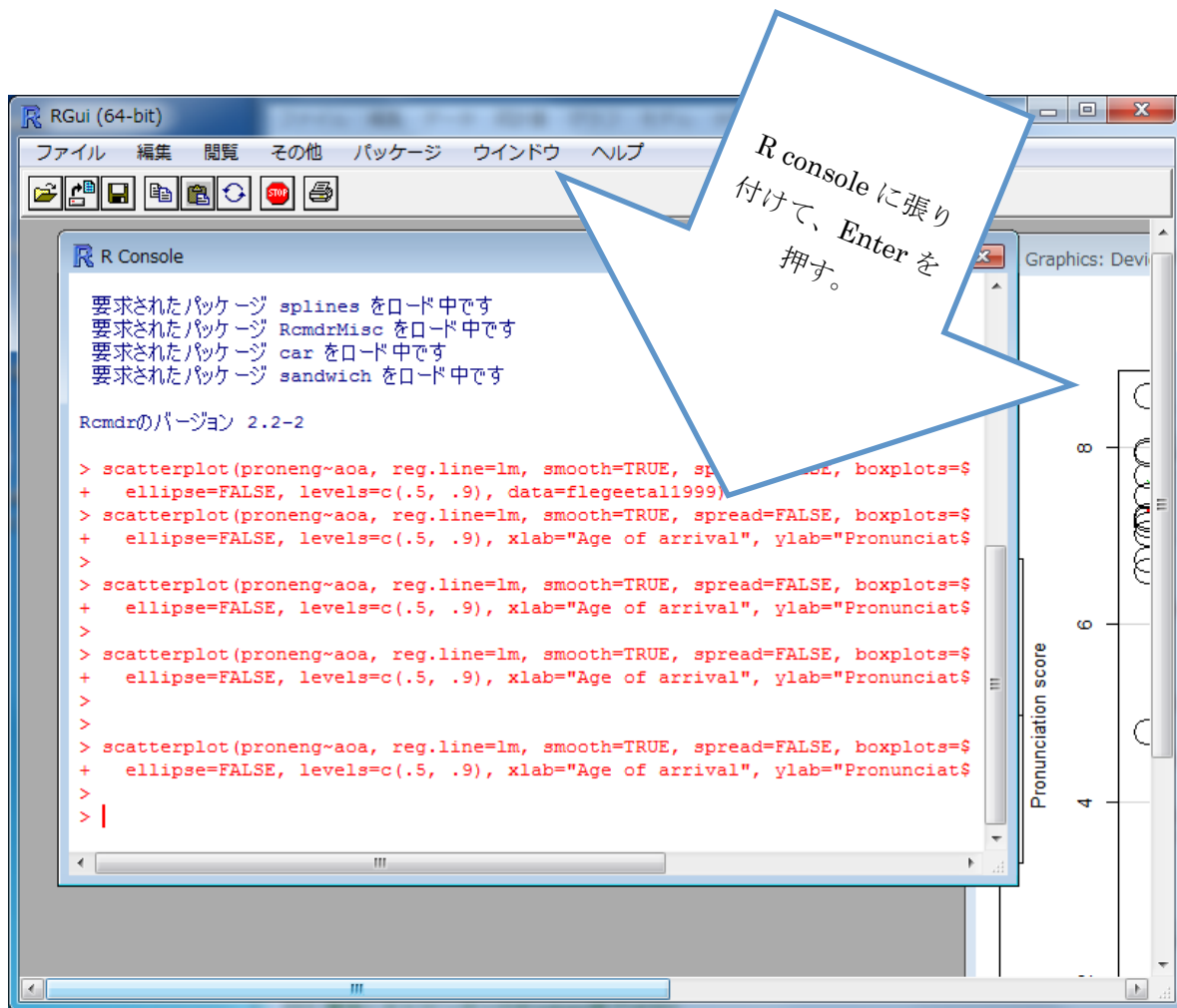
6.1.1 R コンソール内の散布図の修正

The screenshot shows the R Commander window with the following components:

- Menu Bar:** ファイル, 編集, データ, 統計量, グラフ, モデル, 分布, ツール, ヘルプ
- Toolbar:** データセット: flegeetal1999, データセットの編集, データセットを表示, モデル: <アクティブモデルなし>
- Script Editor:** Contains R code for reading the 'flegeetal1999' dataset and creating three scatterplots. The third plot is highlighted in blue.
- Output Console:** Shows the execution of the commands, including the third scatterplot command.
- Message Console:** Displays two messages: a warning about the SDI setting and a note about the number of rows and columns in the dataset.

Callout Box: A blue-bordered box with a white background and a blue arrow pointing to the third scatterplot command in the script editor. It contains the following text:

“Pronunciation score” と “data=flegeetal1999” の間に、“cex=1.6” と入力し、



6.1.2 カテゴリーごとに散布図を観察する

Data: DeKeyser2000.sav

Name: dekeyser

R Cmdr: グラフ > 散布図

R Code:

```
dekeyser$Status<-factor(dekeyser$Status)
```

```
scatterplot(GJTScore~Age | Status, by.groups=TRUE, data=dekeyser)
```

R 散布図

データ

オプション

x 変数 (1つ選択)

age

gjttscore

y 変数 (1つ選択)

age

gjttscore

層別 : status

部分集合の表現

<全ての有効なケース>

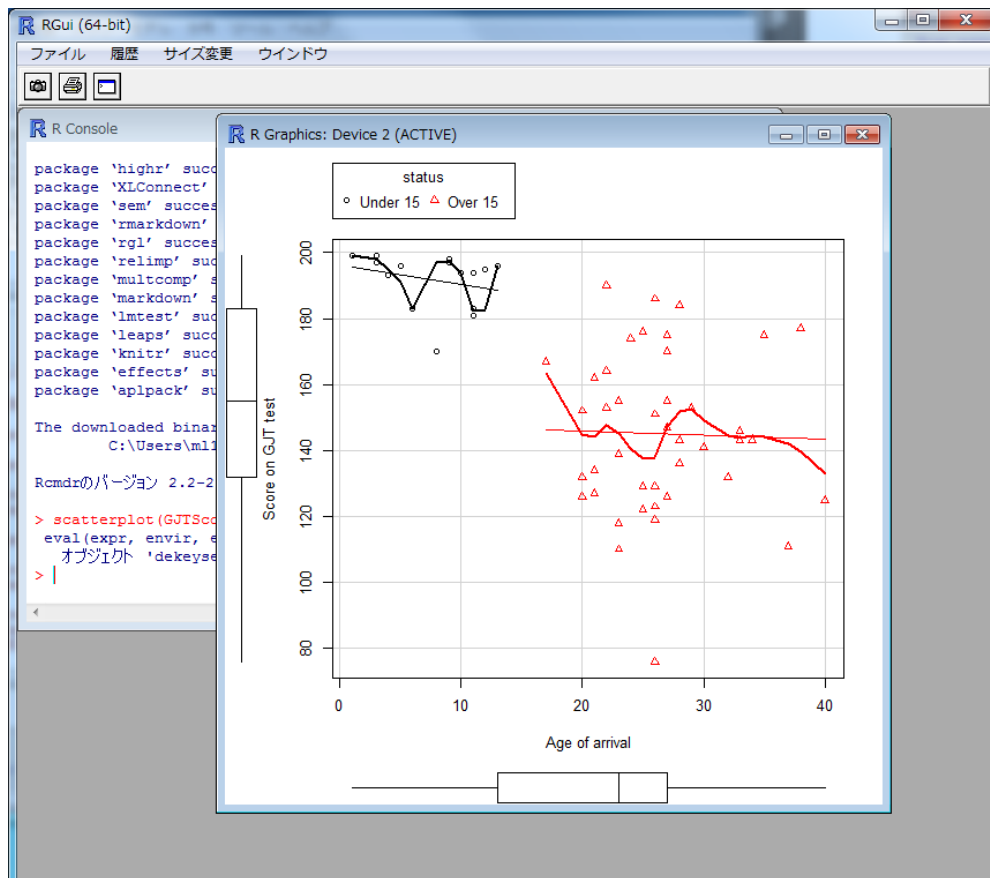
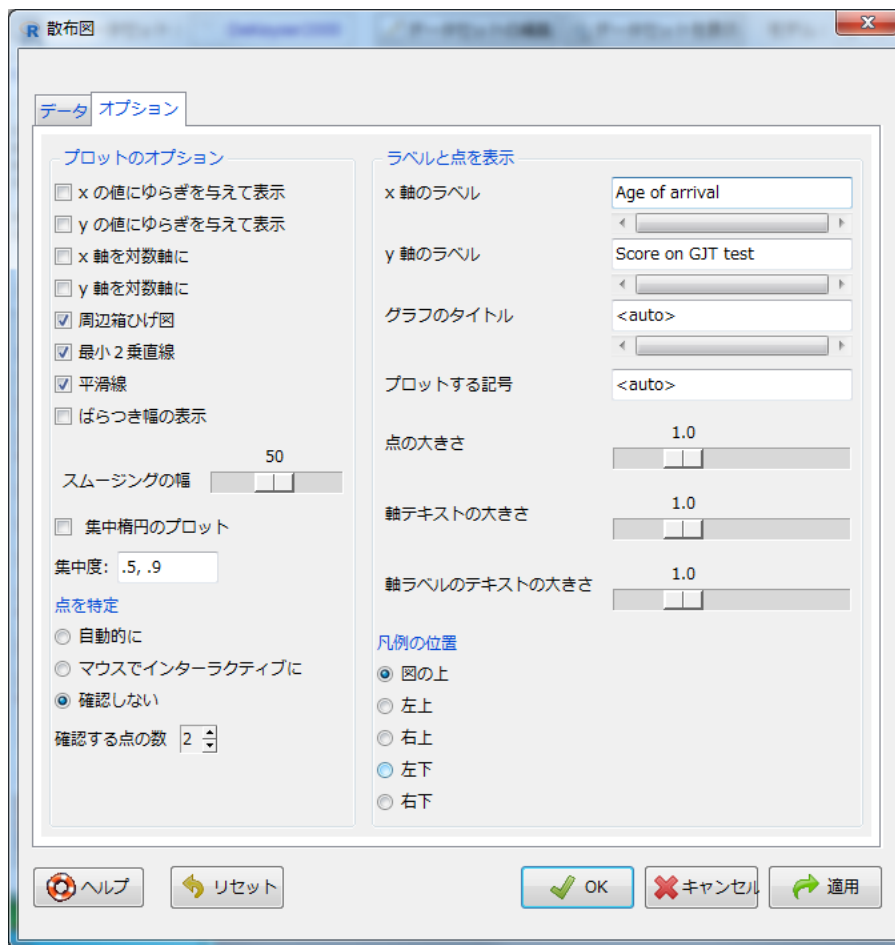
ヘルプ

リセット

OK

キャンセル

適用



6.1.3 複数の散布図

Data: LarsonHall2008.sav

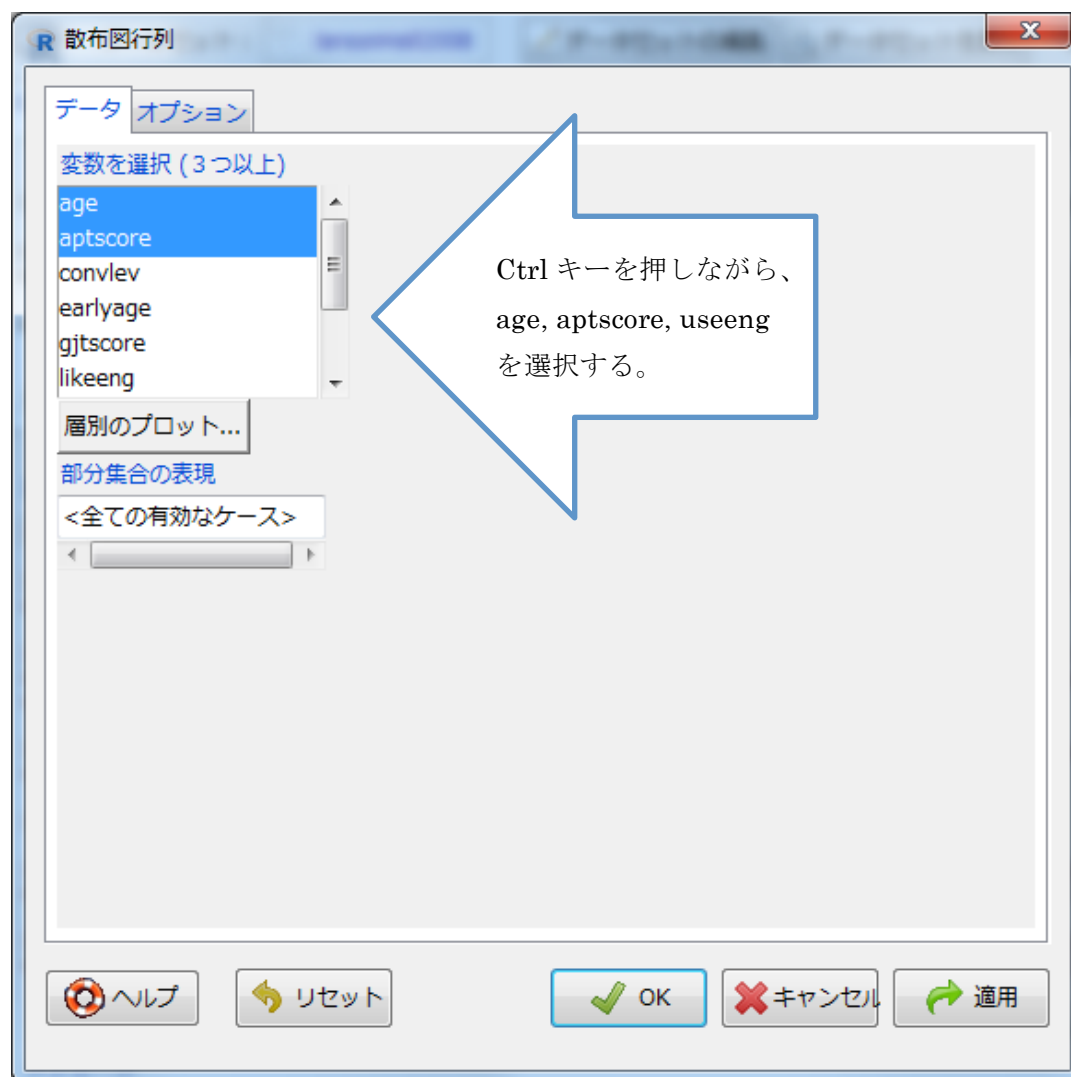
Name: laronHall2008

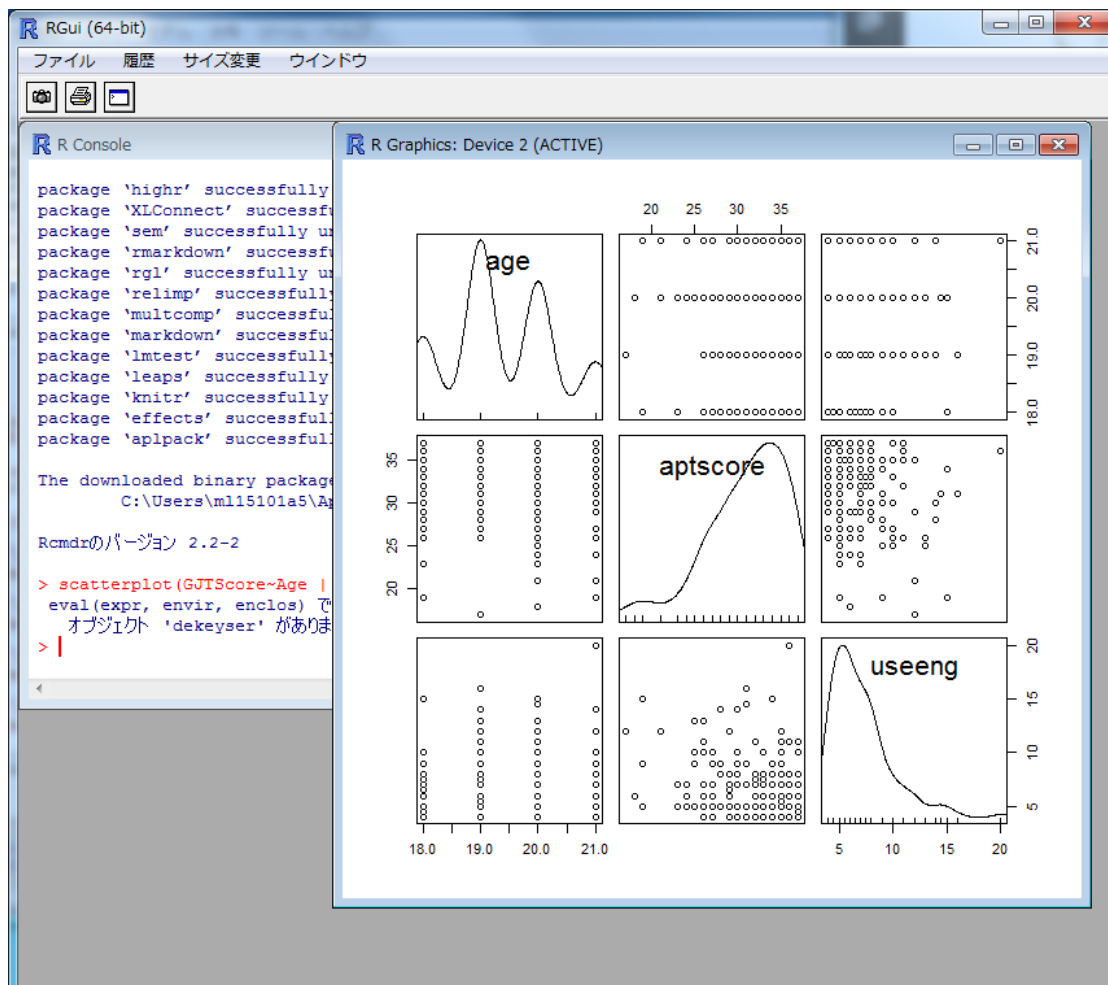
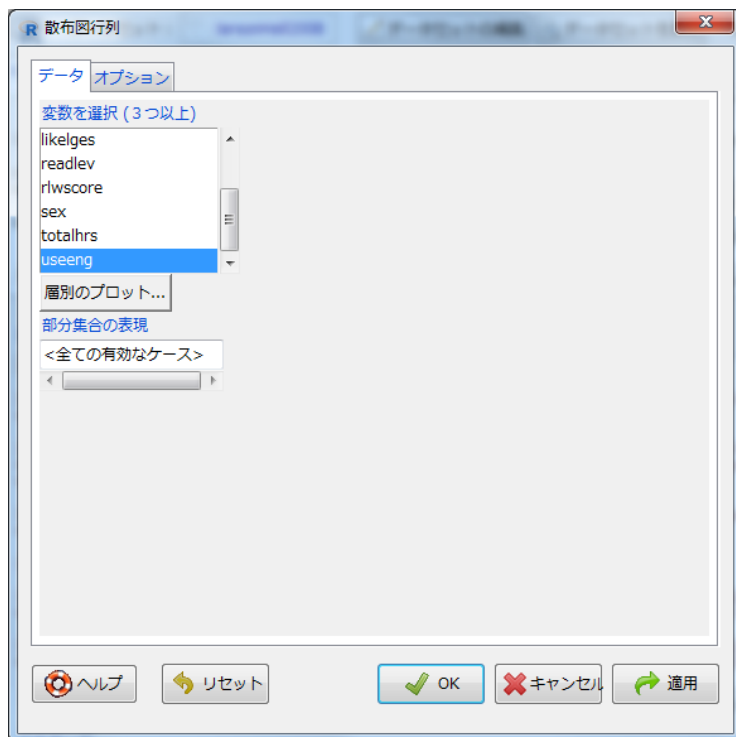
R Cmdr: グラフ > 散布図行列

R Code:

```
library(car)
```

```
scatterplot.matrix(~age+aptscore+useeng, reg.line=lm, smooth=TRUE, span=0.5,  
diagonal = 'density', data=larsonhall2008)
```





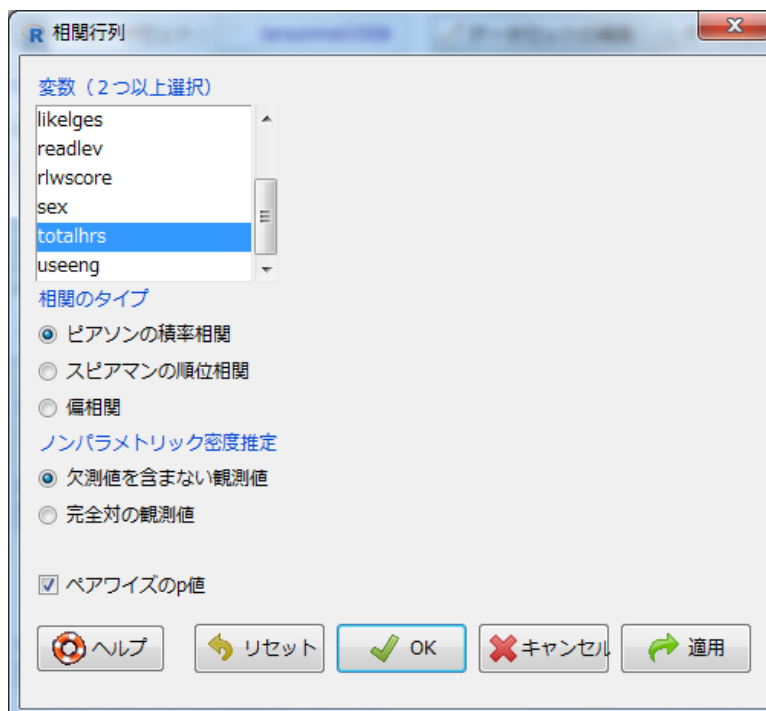
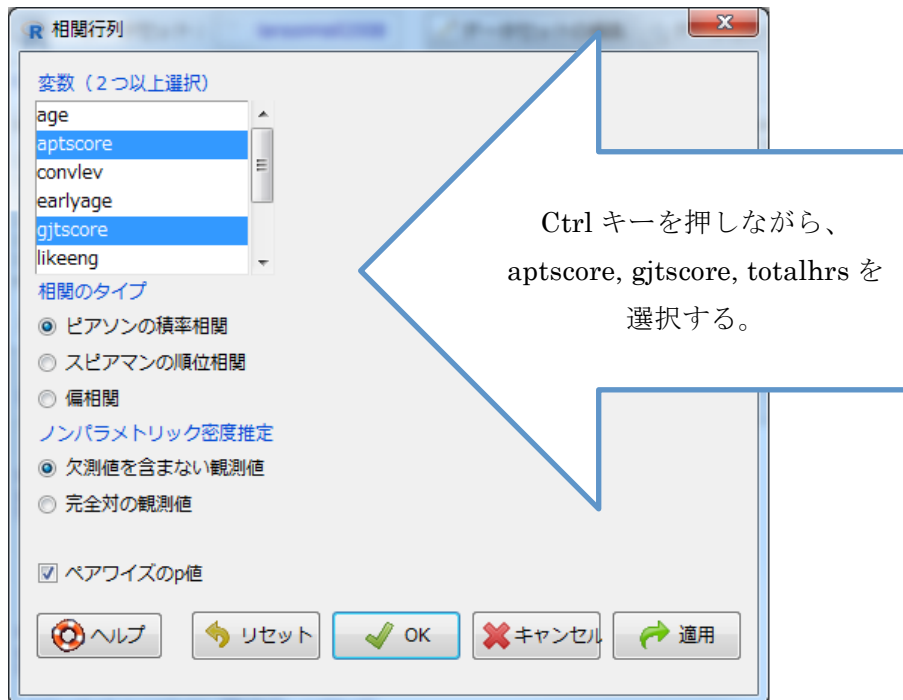
6.3 相関係数の計算

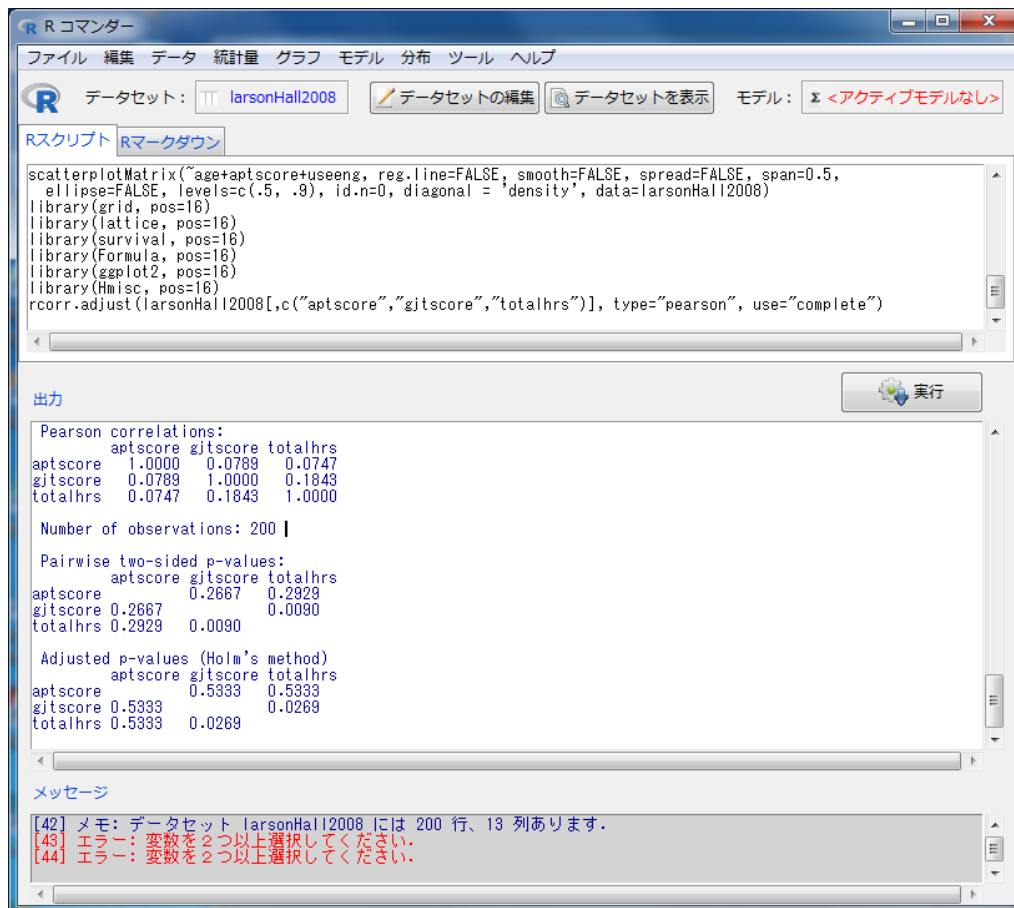
R Cmdr: 統計量>要約>相関行列

R Code: library(Hmisc, pos=4)

```
rcorr.adjust(larsonhall2008[,c("aptscore", "gjtsscore", "totalhrs")],  
type="pearson")
```

```
cor.test(larsonhall2008$gjtsscore, larsonhall2008$aptscore,  
method="kendall")
```





Pearson correlations:

	aptscore	gitscore	totalhrs
aptscore	1.0000	0.0789	0.0747
gitscore	0.0789	1.0000	0.1843
totalhrs	0.0747	0.1843	1.0000

Number of observations: 200

Pairwise two-sided p-values:

	aptscore	gitscore	totalhrs
aptscore		0.2667	0.2929
gitscore	0.2667		0.0090
totalhrs	0.2929	0.0090	

Adjusted p-values (Holm's method)

	aptscore	gitscore	totalhrs
aptscore		0.5333	0.5333
gitscore	0.5333		0.0269
totalhrs	0.5333	0.0269	

6.3.1 ロバスト相関

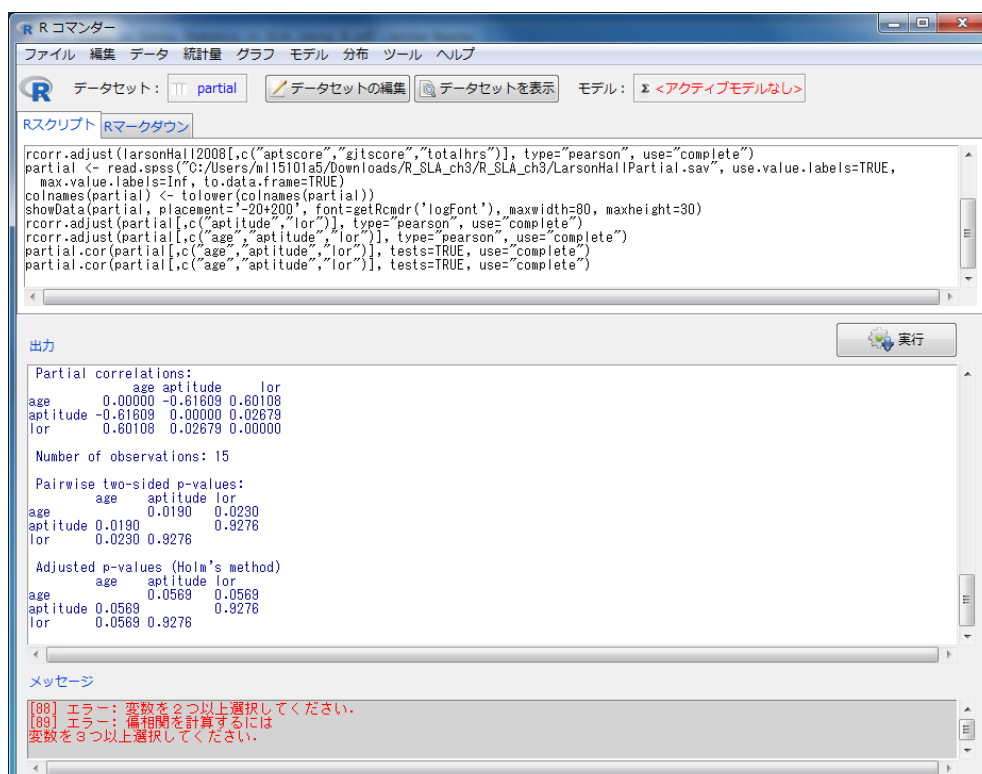
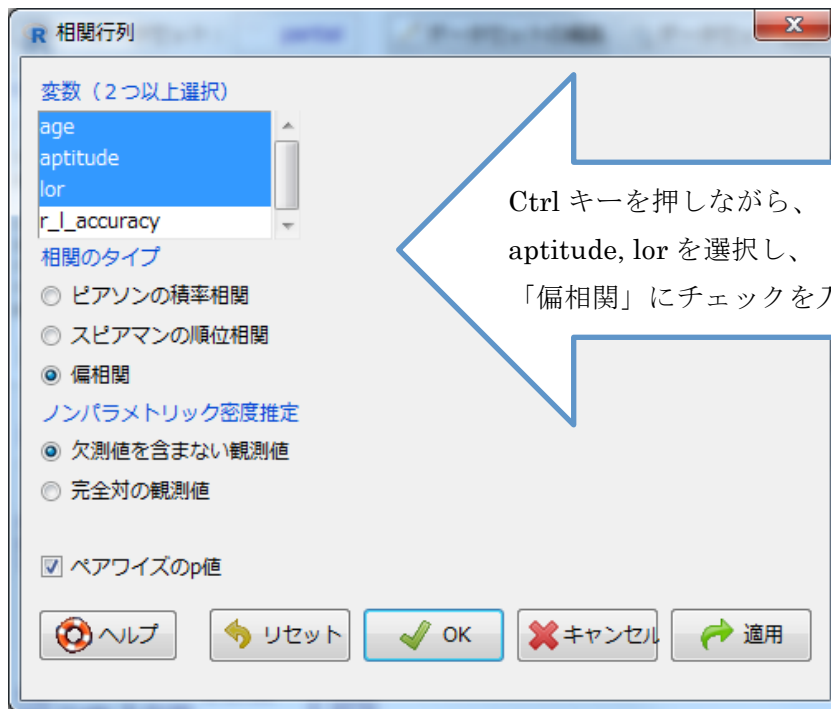
6.5 偏相関

Data: LarsonHallPartial.sav

Name: partial

R Cmdr: 統計量>要約>相関行列

R Code: `partial.cor(partial[,c("age", "LOR", "aptitude")], use="complete.obs")`



Partial correlations:

	age	aptitude	lor
age	0.00000	-0.61609	0.60108
aptitude	-0.61609	0.00000	0.02679
lor	0.60108	0.02679	0.00000

Number of observations: 15

Pairwise two-sided p-values:

	age	aptitude	lor
age		0.0190	0.0230
aptitude	0.0190		0.9276
lor	0.0230	0.9276	

Adjusted p-values (Holm's method)

	age	aptitude	lor
age		0.0569	0.0569
aptitude	0.0569		0.9276
lor	0.0569	0.9276	

6.6 点双列相関係数と評価者間信頼性

点双列相関係数とは、2つの変数のうち、一方の変数が2値しかとらず、もう一方の変数が連続変数の場合の相関係数のこと。

6.6.1 点双列相関係数とテストの解析

Data: LarsonHallGJT.sav

Name: LHtest

R Cmdr: 統計量>次元解析>尺度の信頼性

R Code: reliability(cov(LHtest[,c("Q43","Q44","Q45","TotalScore")],use="complete.obs"))

6.6.2 評価者間信頼性

R Cmdr: 統計量>次元解析>尺度の信頼性

R Code:

```
reliability(cov(MDM[,c("m001","m002","m003","m004","m005","m006","m007",  
"m008","m009","m010")], use="complete.obs"))
```

```
library(psych)
```

```
alpha(LHtest)
```

```
rcorr.adjust(MDM[,c("m001","m002","m003","m004","m005","m006","m007","m008",
```

```
"m009","m010"),type="pearson")
```