αが低いのは、項目数のせい…? (1)

論文を読んでいると、時折(最近は多くなっているようにも感じるが…気のせいか?)「α 係数は.684と若干低めであったが、これは項目数が少ないためと考えられ、信頼性は利用に 十分と判断できる…」といったような、かなり手前味噌な表現を目にする…

確かに、項目間相関が同一であるならば、項目数が少ないほどα係数は小さくなります。 これを知ったのは、確か SPSS の分厚いマニュアルだったと記憶しているのだけど、そうい ばα係数と項目数、項目間相関の関係が示されている一覧表で見たことがない…。そこで今 回はこれを作ってみたいと思います。そんなもの、シミュレーション的にやらずとも、計算 できるという突っ込みは無しでお願いしたいのですが…

では何が必要かと考えてみると、項目間相関が一定の複数項目からなるデータですね。た とえば、すべての項目間相関が .50 になる 10 項目を作成しておいて、そこから3項目をと り出してα係数を求め、4項目を取り出してα係数を求め、5項目取り出してα係数を求め …というようなことを繰り返して、α係数の変化を見てみようと思います。

相関の時と同様に, MASS パッケージに入っている mvrnorm 関数が使えるのですが, 10 項目だと, matrix(c(1, 0.5, 0.5, 0.5 … 1)と, とんでもない数(10×10)を書か なければなりません。これは面倒です…

ということで、これをエクセルを使って作っておき、それを読み込ませるという方法でや ってみようと思います。どうせ簡単にやるならもっと多い項目でもそれほど手間ではないの で、30項目にしましょう。以下のように1と0.5が並んだマトリックスを作成しておいてく ださい。

	A	B	C	D	E	F	G	Н	I	J	K	L	M	N	0	P
1		q1	q2	q3	q4	q5	q6	q7	q 8	q9	q10	q11	q12	q13	q14	q15
2	q1	1	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	1
3	q2	0.5	1	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	1
4	q3	0.5	0.5	1	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	I
5	q4	0.5	0.5	0.5	1	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	1
6	q5	0.5	0.5	0.5	0.5	1	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	1
7	q6	0.5	0.5	0.5	0.5	0.5	1	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	I
8	q7	0.5	0.5	0.5	0.5	0.5	0.5	1	0.5	0.5	0.5	0.5	0.5	0.5	0.5	I
9	q8	0.5	0.5	0.5	0.5	0.5	0.5	0.5	1	0.5	0.5	0.5	0.5	0.5	0.5	I
10	q9	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	1	0.5	0.5	0.5	0.5	0.5	I
11	q10	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	1	0.5	0.5	0.5	0.5	1
12	q11	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	1	0.5	0.5	0.5	1
13	q12	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	1	0.5	0.5	1
14	q13	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	1	0.5	1
15	q14	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	1	1
16	q15	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	
17	q16	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	1
18	q17	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	1
19	q18	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	1
20	a19	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	1

そして csv 形式で保存し(mat30.csv という名前にしておきました), R に読み込みます。 今回は行列ともに名前が入っているので

x <- read.csv("mat30.csv", header=TRUE, row.names = 1, fileEncoding="CP932")

こんな感じでしょう。fileEncoding="CP932"はMac の場合は必要ですね。

次は, MASS パッケージに入っている mvrnorm 関数を使ってデータを発生させます。しかし, mu=の後に平均値0を30個並べるのもまた面倒です。そこで rep(a, b)という関数を使って手間を省きます。これはaをb回繰り返しなさいということです。データ数は、とりあえず200個にしておきましょうか。そしてできたデータをデータフレーム形式に変換しておきます。

library(MASS)

mux <- rep(0, 30)
data <- mvrnorm(n= 200, mu= mux, Sigma= x, empirical= TRUE)
d.data <- data.frame(data)</pre>

データができたら、平均値や標準偏差、そして cor(d.data)を使って指定した通りの相関になっているかどうかを確認してみましょう。

これでデータができたので、後は項目数を指定してαを計算してみます。α係数の算出は psych パッケージですので、これを読み込んでおきます。項目数は、alphaの後のカッコ内 で指定する列数でコントロールすると簡単でしょう。たとえば3項目のαなら以下のようで す(もちろん d. data[10:12]などでも同じになりますが…)。

library(psych) alpha(d.data[1:3])

結果を見ると、α係数は .75 です。つまり、3項目間の平均相関係数が .50 であれば、α 係数は .75 になるということですね。軽く.70 をクリアしています。…ということは、5項 目で .70 に届かないということは、どれくらいの項目間相関なのでしょうかね…

```
> alpha(d.data[1:3])
Reliability analysis
Call: alpha(x = d.data[1:3])
  raw_alpha std.alpha G6(smc) average_r
                                           mean
                                                  sd
      0.75
                0.75
                        0.67
                                   0.5 1.2e-17 0.82
Reliability if an item is dropped:
   raw_alpha std.alpha G6(smc) average_r
q1
        0.67
                  0.67
                           0.5
                                     0.5
q2
        0.67
                  0.67
                           0.5
                                     0.5
q3
        0.67
                  0.67
                           0.5
                                     0.5
Item statistics
         r r.cor r.drop
    n
                             mean sd
q1 200 0.82 0.67
                  0.58 -3.4e-17
                                  1
                    0.58 1.7e-17
q2 200 0.82 0.67
                                   1
                   0.58 3.7e-17
q3 200 0.82 0.67
                                   1
```

もしかすると、先の「3項目間の<u>平均相関係数が</u>.50 であれば」、というところに引っか かった人がいるかもしれません。しかし計算式上では、今回のデータのようにすべての相関 係数が.5 でも、以下のように平均が.5 でも同じ値になります。

	qq1	qq2	qq3
qq1	1	0.3	0.7
qq2	0.3	1	0.5
qq3	0.7	0.5	1

懐疑的な人は、上の表について以下のように確認してみてください。

x0 <- matrix(c(1, 0.3, 0.7, 0.3, 1, 0.5, 0.7, 0.5, 1), ncol=3)
data0 <- mvrnorm(n = 200, mu = c(0, 0, 0), Sigma =x0, empirical = TRUE)
d.data0 <- data.frame(data0)
alpha(d.data0[1:3])</pre>

こちらもα係数は.75になります。

では、使う項目数をいろいろと変えてみて、α係数がどのように変化するのかを確認して みてください。 さて、こういう手作業をやっていると、やはり一気にやりたくなります。2項目から 30 項目までのαをグラフにするとどうなるのだろう…なんてことが気になってきます。

相関の時のやり方をちょっと変えれば簡単にできます。ただし, raw_alpha という数値 は, alpha が出してくるオブジェクトの total の中に入っているので,一度 total を何か に保存し, さらにそこから raw_alpha 引き出すという手順になります。

```
box0 <- rep(NA, 29)
box <- matrix(box0, ncol=1)</pre>
qq1 <- c(2:30)
for(n in 2:30) {
       al <- alpha(d.data[1:n])$total</pre>
       box [n-1,1] <- al$raw_alpha</pre>
       }
plot(box, type="l")
 なお、raw_alpha を一気に引き出すなら、以下のようにすればできます。
box0a <- rep(NA, 29)
boxa <- matrix(box0a, ncol=1)</pre>
qq1 <- c(2:30)
for(n in 2:30) {
       boxa [n-1,1] <- alpha(d.data[1:n])$total$raw_alpha</pre>
       }
plot(boxa, type="l")
```

box に入れられたα係数をグラフにしたら以下のようになります。 項目数が少ない場合は、1項目の増減がα係数に大きく影響していますが、項目数が 10 あたりになってくると、それほど大きな影響は無いようです。



さて、こうなると .5以外の項目間相関の場合はどうなのかをやってみたくなりますね(?)。 一度相関マトリックスを読み込んでおけば、その変更は簡単です。以下のような1行を加 えれば、自由に変更できます。 0.5 を 0.35 に変えるなら…

x[x==0.5] <- 0.35

ただし、これをやるとxが上書きされてしまうので注意してください。別のファイルを作ってやってもよいでしょう。

xx <- x xx[x==0.5] <- 0.35

実際にやってみてください。

もちろんα係数は低くなりますが、グラフを描いてみると、上と似たようなカーブになり ます。 さて, R はグラフを重ね書きすることができます。今回も, 項目間相関が .50 と.35 の場 合のグラフをひとつにすることができます。

.50 の時の結果(α係数)はbox という名前で存在しています。.35 の時の結果は, box.x という名前の入れ物に入れたとしましょう。これを使って,以下のようにすればひとつのグ ラフにしてくれます。

par(family="Osaka") # Mac のみ plot(box, ylim=c(0.5,1), xlab="", ylab="", type="l") par(new = TRUE) plot(box.x, ylim=c(0.5,1), xlab="項目数", ylab="α 係数", type="l", col="red")

1行目はMac用といってもよいでしょう。Winの場合は不要です。

2行目が box を使ってグラフを描く命令,4行目(から5行目にかけて)が box.x を使 ってグラフを描く命令ということはわかると思います。その間にある3行目が上書きをする ためのものです。つまり box を使ったグラフの上に,そのまま box.x を使ってグラフを描 くことになります。

ホントにそのまま2重に描いてくれるので、軸の目盛などが違っているとグチャグチャに なってしまいます。そこで、今回はX軸は同じなので、Y軸だけ目盛をそろえておきます。 ylim=c(0.5,1)という部分がそれなのですが、カッコの中で目盛の最小と最大を指定する だけでOKです。

xlab="", **ylab=""**の部分は, それぞれの軸に対する説明を書き込める部分です。これ は何も指定しないとデフォルトの設定で表示されます。先の図で X 軸には **Index**, Y 軸には **box** と表示されていますが, これがデフォルトの結果です。これをコントロールする部分で す。**""**で囲んだ文字列を表示できるのですが, 2行目は文字列がありません。つまり,何も 表示させないという設定です。上書きする時に,「項目数」と「α係数」を入れるようにして あります。

最後の col="red", そのままの意味で, 赤で描きなさいということです。デフォルトは 黒なはので, 黒の線が .50 の時の結果, 赤の線が .35 の結果です。

これをやってみると、次のような図が出力されます。

やはりカーブの形は似ていますね。しかし、線の差は項目数が少ない時に大きく、項目数 が増えると小さくなる傾向がわかります。



項目数