

α が低いのは、項目数のせい…？ (3)

前回の資料などは、わざわざシミュレーション的にやらずとも、計算でできます。それをなぜデータを発生させてまでやってみたかという、今回のようなことをやってみる前ふり(?)的な位置づけでした。

シミュレーション的に一番やってみたかったのは、今回のようなことです。それは…

先行研究で α 係数が低い尺度を使うと、残念なこと(自分がとったデータでは α が.7に満たないとか…)になる可能性が結構高いことは経験的に知っています。それはサンプルの相関係数が母集団の相関係数と一致しない(ブレる)ことに起因しているはず。[相関係数について]でサンプルと母集団の相関係数とのブレは確認しましたが、それなら α 係数もブレるはず。では、どれくらいブレるのか？

これをシミュレーション的にやってみたかったのです。

たとえば、10000人の母集団における3項目間の項目間相関が.50だったとします。そこから200名を抽出して、 α を計算すると、どの程度の値のブレがでてくるのかを確認したかったわけです。これはたぶん計算ではできないはず…(?)

これをやってみるスクリプトを作ってみます。

```
x <- read.csv("mat30.csv", header=TRUE, row.names = 1,
fileEncoding="CP932")
library(MASS)
library(psych)
mux <- rep(0, 30)
data <- mvrnorm(n= 10000, mu= mux, Sigma= x, empirical= TRUE)
d.data <- data.frame(data)
qq1 <- c(1:10000)
qq2 <- sample(qq1, 200, replace = FALSE)
sub.d.data <- subset(d.data[qq2, ])
alpha(sub.d.data[1:3])
```

こんな感じですかね。

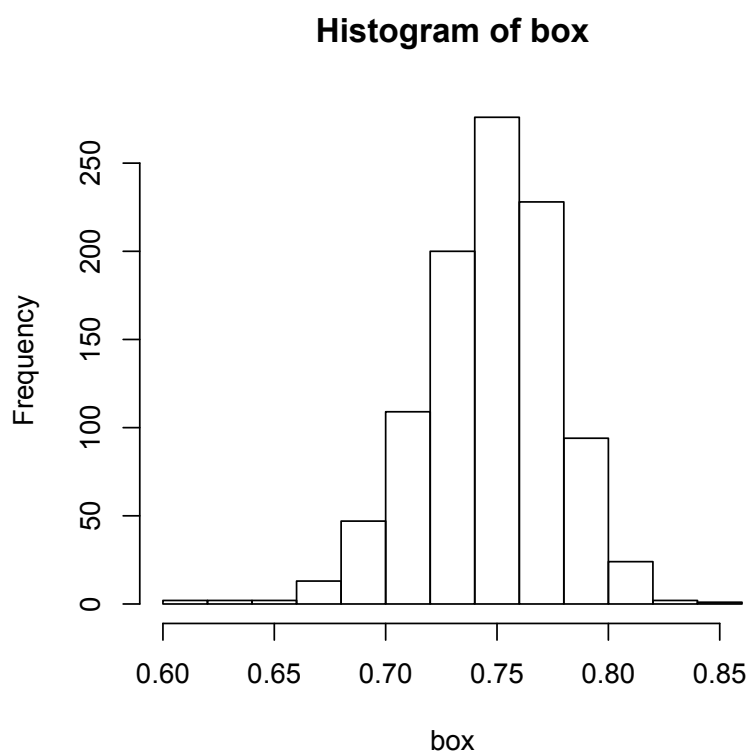
これまでに確認したように、項目間相関の平均が.50の場合、 α 係数は.75になります。つまり母集団における α は.75ということでしょう。サンプリングしたものから計算される α 係数は、ここからどれくらいブレるのか？

何回かこれを繰り返してみると、.7台が連発。想像していたよりブレないか…

ということで、1000回繰り返してみるとどうなるかやってみました。

```
x <- read.csv("mat30.csv", header=TRUE, row.names = 1,
fileEncoding="CP932")
library(MASS)
library(psych)
mux <- rep(0, 30)
data <- mvrnorm(n= 10000, mu= mux, Sigma= x, empirical= TRUE)
d.data <- data.frame(data)
qq1 <- c(1:10000)
box0 <- rep(NA, 1000)
box <- matrix(box0, ncol=1)
for(m in 1:1000) {
  qq2 <- sample(qq1, 200, replace = FALSE)
  sub.d.data <- subset(d.data[qq2, ])
  box[m,1] <- alpha(sub.d.data[1:3])$total$raw_alpha
}
describe(box)
hist(box)
```

平均は .75, 標準偏差 0.03, 最小が .60 で最大が .84。ヒストグラムは以下のよう。



.7 台が連発したのはうなずけるヒストグラムです。私が想像していたほどはブレないようですね。もっとブレると思っていたのに…。相関が低めにもブレるけれども、高めにブレることもあるので相殺されるのかな…

サンプルの相関は、サンプル数によってもブレ方が違うので、いろいろと試してみてください。

いろいろとやってみての極めて個人的感想なのですが、200 名前後以上のサンプルから作成された尺度で、 α が .80 を越えているようなものならかなり安心して使ってもいいかなという感触です（母集団の相関を結構正確に推定できているし、その値を母集団の相関と仮定してのシミュレーションでも、概ね α は .70 はこえてくるため）。

ある研究で作成された尺度について利用を検討しようという際には、 α が .7 をクリアしていても低ければ低いほど、その尺度使ってデータを集めてみたものの α を計算したら .7 に満たなかったということになる可能性を覚悟すべきでしょうね。もちろん、そこに届かないようなものなら使わない、もしくは修正して再検討。作成された際のサンプル数がかなり少ないならば、あくまでも追試気分で使ってみる、くらいのスタンスがよいのかもしれませんが。

項目数うんぬんの議論は、論文を発表するための数値の解釈の問題であり、著者の姿勢の問題でしょう。それを先行研究として読む読者とか、それを参考にする研究者としての自分は、それに対してこのような利用の仕方をしようかなと思いました。