

15日目：1要因分散分析（被験者間）

さて、本日は因子分析の結果から構成された下位尺度得点について、1要因分散分析を行ってみます。やろうとしていることですが、因子分析の結果から構成された下位尺度得点 (**xx**に入っている、**to.f1**, **to.f2**, **to.f3**) に対して、専攻による差があるかどうかを1要因分散分析を用いて検定を行おうということです。

もちろん、まずはそれぞれの群別の平均値、標準偏差などを確認しておきます。検定をやってから、平均値を確認するのは順序が逆です！！

方法は、昨日もやった **describe.by** が便利でしょう。また箱ひげ図も描かせてみましょう。いずれも昨日の「性別」の部分で、「専攻」に変更すればOKです。

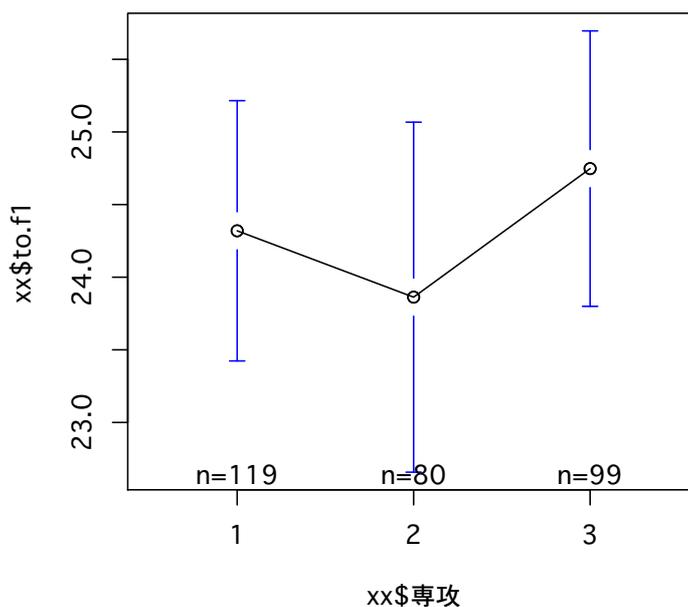
加えて、もう一つ、図を作成してみましょう。箱ひげ図もよいのですが、平均値や標準偏差が出てこないのが難といえれば難です。しかし、こんな図も作れます。

これは各群の平均値を折れ線でつなぎ、それぞれの平均値から1標準偏差分の幅も同時に示してくれています。つまりこの図から、3群の重なりはかなり大きいであろうことが読み取れます。

このような図の作成は、**gplots** というパッケージを使いますので、まずはこれをとってきておく必要があります。ただし、このパッケージは複数の他のパッケージを同時に必要と

するようで、**library(gplots)** などとしてロードしようとする、あのパッケージがない、このパッケージがないと、やたらと警告を出してきます。さらには、パッケージがより新しいバージョンのRにしか対応していないとかで、R本体もインストールすることに…。そんなこんなで警告されたパッケージを4つか5つかとってきておくと、やっとロードできるようになりました（私の場合だけ？）。

ロードできてしまうと後は簡単で、以下のように命令すれば描いてくれます。ちなみに、「専攻」という日本語を使うので、**par** でフォントの設定を先におきます。



```
par(family="Osaka")  
plotmeans(xx$to.f1 ~ xx$専攻)
```

まずはこのようにして、データをざっと眺めてみましょう。

そして `describe.by` 出力結果を整理すると、右のような状況になります。このグループ間の差について検定をするわけです。

		group:1	group:2	group:3
to.f1	mean	24.32	23.86	24.75
	sd	4.94	5.41	4.75
	n	119	80	99
to.f2	mean	13.58	12.38	13.76
	sd	3.77	3.85	3.9
	n	118	80	99
to.f3	mean	7.58	8.3	8.51
	sd	2.96	3.38	3.29
	n	119	80	99

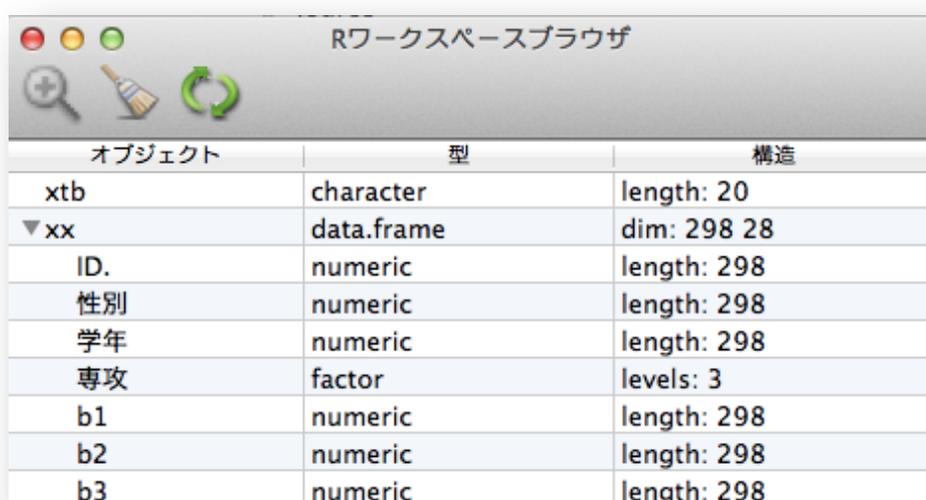
さて、この1要因分散分析にはいくつか留意点があります。まず一つ目ですが、独立変数を `factor` 型とよばれる形式にしておく必要があります。これをしておかないと、間違った計算結果を返してくるようです。

今回の独立変数は「専攻」ですが、これに対して…

```
xx$専攻 <- as.factor(xx$専攻)
```

これで `factor` 型に変更されます。

現在どのような型になっているかの確認は、「ワークスペース」から「ワークスペースブラウザ」を選びます。すると、「R ワークスペースブラウザ」というウインドが開くので、`xx`の前の横向き▲マークをクリックして、中を一覧します。`factor` 型に変更されていたなら、図のように型のところが `factor` になっています。



The screenshot shows the 'Rワークスペースブラウザ' window. It displays a table with three columns: 'オブジェクト' (Object), '型' (Type), and '構造' (Structure). The 'xx' object is expanded to show its components:

オブジェクト	型	構造
xtb	character	length: 20
▼ xx	data.frame	dim: 298 28
ID.	numeric	length: 298
性別	numeric	length: 298
学年	numeric	length: 298
専攻	factor	levels: 3
b1	numeric	length: 298
b2	numeric	length: 298
b3	numeric	length: 298

別の方法としては、以下のような命令を実行してみます。

class(xx\$専攻)

factor型に変更されていたなら、Factorと出力されます。

間違った計算結果を出さないためにも、この変更と確認は大事です。

もう一つは、等分散の確認が重要になります。等分散性を仮定する場合と、仮定しない場合で使う命令が違ってくるので、まずはこれをチェックしておきます。

bartlett.test(to.f1 ~ 専攻, xx)

バートレット検定というものですが、等分散性の検定です。帰無仮説が「等分散である」なので、棄却されなければ等分散性を仮定する場合、棄却されたら等分散性を仮定しない場合となります。

結果は以下のように返されます。Bartlett's K-squaredは、 χ^2 統計量です。この結果、to.f1は等分散性を仮定する方法でOKということになります。今回は、3つの下位尺度すべてで等分散を仮定する方法でOKです。

```
> bartlett.test(to.f1 ~専攻, xx)

Bartlett test of homogeneity of variances

data:  to.f1 by 専攻
Bartlett's K-squared = 1.568, df = 2, p-value = 0.4566
```

Rには1要因分散分析を行う命令が3種類ほどあるようですが、今回はaovを使ってみます。使い方は以下のような感じです。カッコの中は、(従属変数「~」 独立変数「,」 データ)と並びます。

```
aov.f1 <- aov(to.f1 ~ 専攻, xx)
```

```
summary(aov.f1)
```

すると右のように結果が返されます。

F(2,295)=0.692, ns

という結果ですね。

3つの下位尺度得

```
> aov.f1 <- aov(to.f1 ~専攻, xx)
> summary(aov.f1)
          Df Sum Sq Mean Sq F value Pr(>F)
専攻      2     35    17.36   0.692  0.501
Residuals 295   7404    25.10
```

点についてすべてやってみると、`to.f2` で5%水準の有意差が、`to.f3` で有意傾向が認められました。そこで`to.f2` についてテューキー法の多重比較を行ってみます。`TukeyHSD()` という形式になるのですが、カッコの中は先の`aov(to.f2 ~ 専攻, xx)`の結果を代入したもの (`aov.f2`) を入れます。もちろん、結果を導く `aov(to.f2 ~ 専攻, xx)`をそのまま入れてもOKです。

TukeyHSD(aov.f2)

もしくは

TukeyHSD(aov(to.f2 ~ 専攻, xx))

結果は右のようです。
`group:2` と `group:3` の間に5%水準での有意差が認められています。

```
> TukeyHSD(aov(to.f2 ~ 専攻, xx))
Tukey multiple comparisons of means
 95% family-wise confidence level

Fit: aov(formula = to.f2 ~ 専攻, data = xx)

$専攻
      diff      lwr      upr    p adj
2-1 -1.209746 -2.5183410 0.09884951 0.0767118
3-1  0.172830 -1.0586575 1.40431751 0.9415471
3-2  1.382576  0.0241914 2.74096012 0.0449963
```

結果を簡単にまとめてしまえば、以下の表のようになるでしょう。

		group:1	group:2	group:3	分散分析結果
to.f1	mean	24.32	23.86	24.75	ns
	sd	4.94	5.41	4.75	
	n	119	80	99	
to.f2	mean	13.58	12.38	13.76	*
	sd	3.77	3.85	3.9	2<3
	n	118	80	99	
to.f3	mean	7.58	8.3	8.51	ns
	sd	2.96	3.38	3.29	
	n	119	80	99	

* p<.05; * p<.01

等分散性を仮定しない場合や、シェッフエやボンフェローニといった多重比較もできますが、それはそれぞれで調べてください。

本日はここまでです。