

16日目：2要因分散分析（被験者間）

さて、本日は2要因分散分析を行ってみます。一昨日、昨日という流れ（分析としての）からするとおかしいことにはなりますが、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**）について、性と専攻を要因とする2要因分散分析を行うことをやってみます。

しつこいですが、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…。ですが、今回は群を分ける変数が2つになります。**describe.by** は、複数の群を分ける変数に対応してくれるのか？

結論から言うと、やってくれます。**describe.by(xx[26:28], xx[3:4])** というような命令をやってくれます。しかし、「性別」と「専攻」は2列目と4列目であり、3列目に余分なものが入っているので、**xx[2:4]** とやるとおかしなことになってしまいます。

ここで、「変数をまとめる」ということができたことを思い出してください。まとめてしまえば、一気に指定できます。つまり…

```
s.s <- c("性別", "専攻")
describe.by(xx[26:28], xx[s.s])
```

これで、性別、専攻別に計算してくれます。

ちなみにですが、**describe.by** では(**xx[26:28]**, **xx[s.s]**, **mat=TRUE**) というように **mat=TRUE** を加えることもできます（デフォルトは、**mat=FALSE** になっています）。結果は同じなのですが、出力の形式（順）が少し違っています。自分が作りたい形式に便利な方を選んでおくと、後が楽でしょう。

to.f1 の結果を簡単にまとめる
と右のような表になります。

		専攻			
		1	2	3	
性別	1	mean	23.45	23.93	23.35
		sd	5.07	5.66	4.87
		n	29	59	37
	2	mean	24.60	23.67	25.58
		sd	4.89	4.78	4.52
		n	90	21	62

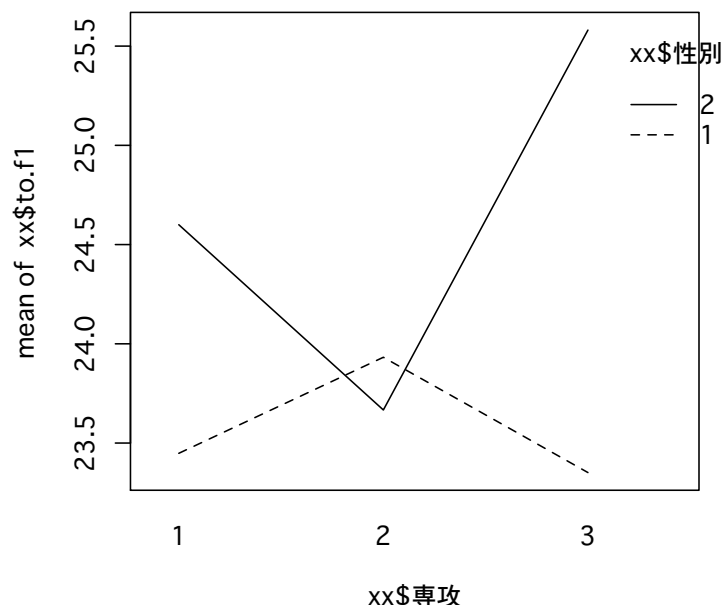
次に、これをグラフにしてみます。2要因分散分析だと、おなじみなのは、あの交互作用を説明する時のグラフでしょう。あのようなものが作れます。

```
par(family="Osaka")
interaction.plot(xx$専攻, xx$性別, xx$to.f1)
```

これで右のようなグラフを作成してくれます。

「性別」という日本語が入るので、フォントの設定をします。`interaction.plot` のカッコの中には、グラフの横軸（独立変数の1つ）、線の区別（独立変数のもう1つ）、グラフの縦軸（従属変数）の順になります。

グラフからだと、交互作用がありそうにも思えますが、実際はどうなのでしょう…



さて、ここから分析に入りますが、1要因分散分析同様、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。

今回の独立変数は「性別」と「専攻」の2つですから、これに対して…

```
xx$性別 <- as.factor(xx$性別)
xx$専攻 <- as.factor(xx$専攻)
```

これで2つを `factor` 型に変更します。確認の仕方は昨日と同じです。

Rには2要因分散分析を行う命令が2種類ほどあるようですが、今回も `aov` を使ってみます。使い方は以下のようなようです。カッコの中は、（従属変数「~」独立変数1「*」独立変数2「,」データ）と並びます。

```
aov2.f1 <- aov(to.f1 ~ 性別 * 専攻, xx)
summary(aov2.f1)
```

結果は、次の図のように表示されます。典型的な分散分析表ですので、特に説明は不要ですね。

性別の主効果が、 $F(1,292)=4.135$ で5%水準の有意。専攻の主効果、交互作用は有意ではないということが読み取れます。先に作成したグラフでは相互作用がありそうな感じもありましたが、検定の結果、性差のみが有意ということがわかりました。

```
> aov2.f1 <- aov(to.f1 ~ 性別 * 専攻, xx)
> summary(aov2.f1)
          Df Sum Sq Mean Sq F value Pr(>F)
性別      1    103  102.79   4.135 0.0429 *
専攻      2     19   9.71    0.391 0.6770
性別:専攻  2     58  28.93   1.164 0.3137
Residuals 292   7259   24.86
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> |
```

私がやるなら、このような結果で十分なのですが…。必要な場合もあるでしょうから、とりあえず多重比較のやり方を書いておきます。

チューキー法なら、昨日の命令を少し変えるだけでできます。TukeyHSD()のカッコの中を、本日のもの書き換えればそれでOKです。

TukeyHSD(aov2.f1)

もしくは

TukeyHSD(aov(to.f1 ~ 性別 * 専攻, xx))

もちろん、単純主効果を計算することもできますが、ここでは省略します。

私自身が分散分析はかなり苦手ということもあり、十分な説明はできていませんが、詳しいwebもありますし、便利そうな自作関数も多く公開されているので、必要ならばそちらをあたってください。

本日はここまでです。