15 日目：１要因分散分析（被験者間）

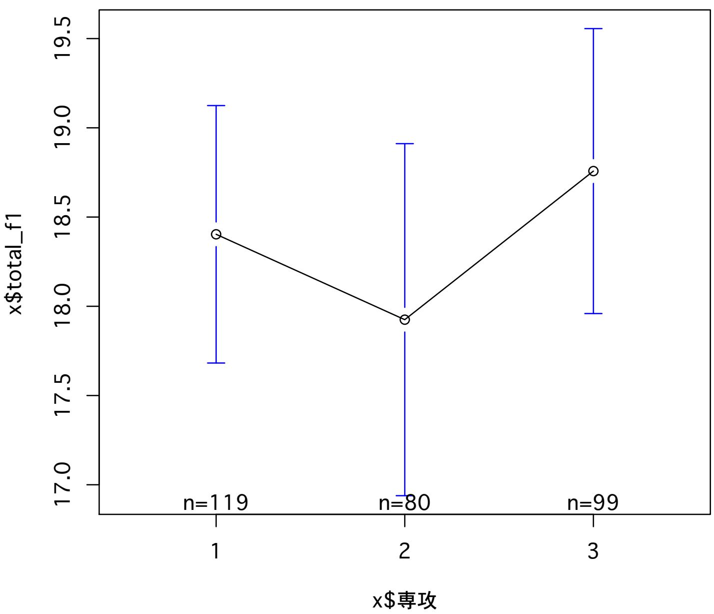
さて，本日は因子分析の結果から構成された下位尺度得点について，１要因分散分析を行ってみます。やろうとしていることですが，因子分析の結果から構成された下位尺度得点（total_f1，total_f2，total_f3）に対して，専攻による差があるかどうかを１要因分散分析を用いて検定を行おうということです。

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 describeBy が便利でしょう。もちろん，ヒストグラムも描かせおきましょう。

加えて，もう一つ，図を作成してみましょう。ヒストグラムは平均値や標準偏差が出てこないのが難といえば難です。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 95%信頼区間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ことが必要です。もちろん，library (gplots)で読み込んでおきます。

命令は簡単で，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しておきます。これをしておかないと，（たぶん）文字化けします。




```
par(family="Osaka")
plotmeans(x$total_f1 ~ x$専攻)
```

まずは describeBy で算出した基礎統計量，ヒストグラム，そして先のグラフなどを使って，データをじっくり眺めておきましょう。

なお describeBy 出力結果を整理すると，右のような状況になります。

			group1	group2	group3
total_f1	機能性	平均値	18.40	17.93	18.76
		標準偏差	3.97	4.43	4.00
		標準誤差	0.36	0.50	0.40
		n	119	80	99
total_f2	デザイン性	平均値	13.58	12.38	13.76
		標準偏差	3.77	3.85	3.90
		標準誤差	0.35	0.43	0.39
		n	118	80	99
total_f3	グレード感	平均値	7.58	8.30	8.51
		標準偏差	2.96	3.38	3.29
		標準誤差	0.27	0.38	0.33
		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，求めているものとは違った計算結果を返してきます。

今回の独立変数は「専攻」ですが，これに対して…

```
x$専攻 <- as.factor(x$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，いくつかの方法があります。ひとつは，以下のような命令を実行することです。

```
class(x$専攻)
```

factor 型に変更されていたなら， Factor と出力されます。

別には,「ワークスペース」から確認する方法があります。メニューから「ワークスペース」→「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，x 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。

Rワークスペースブラウザ

オブジェクト	型	構造
label_f2	character	length: 5
label_f3	character	length: 4
▼x	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

間違った計算結果を出さないためにも，この変更と確認は極めて重要です。

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これをチェックしておきます。

```
bartlett.test(total_f1 ~ 専攻, x)
```

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, 帰無仮説は棄却されないので total_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(total_f1 ~ 専攻, x)

	Bartlett test of homogeneity of variances

data:  total_f1 by 専攻
Bartlett's K-squared = 1.3375, df = 2, p-value = 0.5124
```

そして１要因分散分析ですが，Ｒには１要因分散分析を行う命令が複数あります。今回は aov を使ってみます。使い方は以下のようです。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(total_f1 ~ 専攻, x)
summary(aov.f1)
```

すると右のように結果が返されます。$F(2,295)=0.908$ で，p は 0.404，つまり有意ではないという結果ですね。

```
> aov.f1 <- aov(total_f1 ~ 専攻, x)
> summary(aov.f1)
             Df Sum Sq Mean Sq F value Pr(>F)
専攻          2     31   15.35   0.908  0.404
Residuals   295   4984   16.90
```

３つの下位尺度得点についてすべてやってみると，total_f2 で５％水準の有意差が，total_f3 で有意傾向が認められました。そこで total_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１要因分散分析 aov(total_f2 ~ 専攻, x)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(total_f2 ~ 専攻, x)をそのまま入れても OK です。

```
TukeyHSD(aov.f2)
```

もしくは

```
TukeyHSD(aov(total_f2 ~ 専攻, x))
```

結果は次のようです。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（3-2 の欄で，p adj が 0.05 を下回っています）。diff などの意味は自分で調べてください。

```
> TukeyHSD(aov(total_f2 ~ 専攻, x))
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = total_f2 ~ 専攻, data = x)

$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。

			group1	group2	group3	分散分析結果
total_f1	機能性	平均値	18.40	17.93	18.76	ns
		標準偏差	3.97	4.43	4.00	
		標準誤差	0.36	0.50	0.40	
		n	119	80	99	
total_f2	デザイン性	平均値	13.58	12.38	13.76	*
		標準偏差	3.77	3.85	3.90	2<3
		標準誤差	0.35	0.43	0.39	
		n	118	80	99	
total_f3	グレード感	平均値	7.58	8.30	8.51	ns
		標準偏差	2.96	3.38	3.29	
		標準誤差	0.27	0.38	0.33	
		n	119	80	99	

＊ p<.05

さて，今回の注意点は何といっても独立変数を factor 型にしておくところです。ここまでやってきているならば，`x$専攻 <- as.numeric(x$専攻)`として，factor 型を numeric 型に戻してから，再度，等分散の検定，分散分析を行ってみてください。等分散の検定は同じ結果になるでしょうが，１要因分散分析の結果は変わります。そして，警告やエラーがでません。そのため，間違いに気付きにくいかもしれません。この結果からミスに気付くには，自由度をチェックする必要があります。numeric 型のままだと，自由度が，1,と 296（total_f1 の場合）になります。3 水準の分析なので，分子の自由度が 1 になっているのは，あきらかにおかしいです。十分に気をつけてください。

等分散性を仮定しない場合や，シェッフェ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。