

(6)分散分析

多群(3群以上)のデータに対して、平均や分散に差があるかどうかという検定手法です。
t検定を3群以上に拡張した手法として「一元配置分散分析(もしくは単に分散分析)」があります。
SASで実行するには、GLM(General Linear Model:一般線形モデル)プロシジャを使用します。

◆帰無仮説

すべての群の母平均が等しい

◆対立仮説

母平均が異なる群が少なくとも1つは存在する

※どの群とどの群の間に差があるかを判別することはできません。

プログラム

* サンプルデータ;

```
data work.sample;
  input Group $ Y @@;
  cards ;
A0 143 A0 150 A0 151 A0 152 A0 159 A0 157 A0 152 A0 155
A0 149 A0 155 A1 152 A1 150 A1 159 A1 155 A1 152 A1 156
A1 155 A1 141 A1 158 A1 143 A2 150 A2 142 A2 135 A2 151
A2 144 A2 149 A2 152 A2 148 A2 133 A2 159 A3 132 A3 132
A3 143 A3 140 A3 141 A3 139 A3 141 A3 140 A3 141 A3 151
;
run ;
```

* GLMプロシジャによる分散分析;

```
proc glm data = work.sample;
  class Group ; /* 群(水準)を表す変数を指定 */
  model Y = Group ; /* 左辺に特性値(測定値),右辺に因子(群)をそれぞれ変数で指定 */
quit ;
```

結果

分散分析の結果として、SASのアウトプットウィンドウに分散分析表が出力されます。

変動因	自由度	平方和	平均平方	F値	Pr > F
Model	3	1017.6750000	339.225000	9.11	0.0001
Error	36	1341.1000000	37.252778		
Corrected Total	39	2358.775000			

変動因の[Model][Error][Corrected Total]がそれぞれ[群間][群内][全体]に相当します。
今回の例では、分散分析のp値は0.0001となり、有意水準5%で群間に有意な差があるといえます。
しかし、どの群とどの群の間に有意な差があるかは判断できません。

R2 乗	変動係数	誤差の標準偏差	Y の平均
0.431442	4.133066	6.103505	147.6750

R2乗 : 0~1の値を取り、1に近ければ近いほど、モデルの当てはまり具合が良い。
変動係数 : 誤差の標準偏差が平均の何パーセントを占めるかが示されます。
誤差の標準偏差 : 群内の平均平方の平方根を取った値となります。
Yの平均 : 全群を合わせたときの平均値となります。

※SASは、米国及び各国での米国SAS Institute Inc.の登録商標です。