共分散分析とその応用

2000MM013 平川　晃弘
指導教員　 松田　眞一

1　はじめに

2 群間の平均値の差を検定する場合、様々な条件はあるものの一般的には t 検定が用いられる。また、多群間の平均値の差を検定する場合には分散分析が用いられる。しかし、これらの検定で有意差が認められたからと言って、群間に差があると決めてしまうのは問題がある。なぜなら各群の観測値には共変量が何らかの形で影響していることがあるからである。この共変量が観測値 (変量) に影響を及ぼし、群間に差をもたらした可能性を否定できない。つまり、単純に t 検定や分散分析を用いることは、間違った結論を導いてしまう可能性がある。よって、出来るだけ共変量の影響を取り除き、各群の本当の差の有無を見出すことが必要である。このような問題を解決するために用いられる統計的方法が共分散分析である。

2　共分散分析の成立条件

共分散分析はある条件の下でのみ、適応できる手法であり、その条件は 2 つある。

1.各群の回帰係数が等しい。(回帰の平行性の検定)

2.各群の回帰係数が 0 でない。(回帰の有意性の検定)

3　共変量が 1 つで、2 群の場合の共分散分析

共変量が 1 つならば、各群の回帰直線の大小を比較することになり、次のような段階を経て分析する。(Armitage,Berry[5])

3.1　回帰直線の平行性の検定

まず、次のような回帰直線を考える

$$\text{第 1 群}：y = a_1 + b_1 x$$

$$\text{第 2 群}：y = a_2 + b_2 x$$

この 2 つの回帰直線の傾きが異なると、両群の差を比較できない。そこで、次の検定を行う。

$$H_0(\text{帰無仮説})：b_1 = b_2$$

$$H_1(\text{対立仮説})：b_1 \neq b_2$$

ここで、帰無仮説が棄却されなければ、両群の傾きが等しいと仮定できる。

3.2　回帰直線の有意性の検定

両群の傾きが等しいと仮定できた場合、共通の傾きを b とし、次の検定を行う。

$$H_0(\text{帰無仮説})：b = 0$$

$$H_1(\text{対立仮説})：b \neq 0$$

ここで、帰無仮説が棄却されれば、共通の傾きは 0 でないと仮定でき、共分散分析を適応できる。

3.3　共分散分析

共分散分析は両群の切片が等しいか否かを判断するものであり、次の検定を行う。

$$H_0(\text{帰無仮説})：a_1 = a_2$$

$$H_1(\text{対立仮説})：a_1 \neq a_2$$

ここで、帰無仮説が棄却されなければ、共変量の影響を取り除いた両群に差があるとは言えないことになる。一方で，帰無仮説が棄却されれば、差があると言える。

4　共変量が 2 つで、3 群の場合の共分散分析

共変量が複数の場合の共分散分析は重回帰を用いることになる。分析の流れとしては 3 節とほぼ同様である。(奥野・久米・芳賀・吉沢 [2])

4.1　回帰平面の平行性の検定

まず、次のような回帰平面を考える

$$\text{第 } k \text{ 群}：y_k = a_k + b_{1k} x_1 + b_{2k} x_2 (k = 1, 2, 3)$$

この 3 つの回帰平面の傾きが異なると、各群の差を比較できない。そこで、次の検定を行う。

$$H_0(\text{帰無仮説}) : b_{n1} = b_{n2} = b_{n3} (n = 1, 2)$$
$$H_1(\text{対立仮説}) : H_0\text{でない。}$$

ここで、帰無仮説が棄却されなければ、各群の傾きが等しいと仮定できる。

4.2　回帰平面の有意性の検定

各群の傾きが等しいと仮定できた場合、共通の傾きを b_1, b_2 とし、次の検定を行う。

$$H_0(\text{帰無仮説})：b_1 = b_2 = 0$$

$$H_1(\text{対立仮説})：H_0\text{でない。}$$

ここで、帰無仮説が棄却されれば、共通の傾きは 0 でないと仮定できる。しかし、この検定では b_1, b_2 の一方が 0 でない、もしくは両方が 0 でないということしか分からない。そこで、次の検定を行う。

$$H_0(\text{帰無仮説})：b_n = 0 (n = 1, 2)$$

$$H_1(\text{対立仮説})：b_n \neq 0 (n = 1, 2)$$

ここで、それぞれの帰無仮説が棄却されれば、b_1, b_2 は変量に個別に寄与していると言える。一方で、帰無仮説が棄却された場合、その共変量を除去した上で、共分散分析を行うことになる。

4.3 共分散分析

上記の検定で平行性及び有意性が仮定できるなら、重回帰を用いて分析することに意味があると言える。そこで、次の検定を行う。

$$H_0(\text{帰無仮説}):a_1 = a_2 = a_3$$
$$H_1(\text{対立仮説}):H_0\text{でない。}$$

ここで、帰無仮説が棄却されなければ、共変量の影響を取り除いた各群に差があるとは言えないことになる。一方で，帰無仮説が棄却されれば、差があると言える。

5 変数選択法を用いた共分散分析

共変量が複数の場合の回帰の有意性の検定は、個々の共変量が変量に影響しているか否かは分からなかったために、個別に t 検定を行い、各共変量が変量に寄与しているか否かを判断した。しかし、この検定では「関係しているものを選ぶ」のではなく、「関係のないものを外す」と言うことを行っているに過ぎない。つまり、有意となった共変量が変量に影響を与えていると結論付けてしまうのは問題がある。このような問題を解決するために、共変量の選択方法として、個々の t 検定とは別に、変数選択法を用いてその効果を検証する。

ここでは低出生体重児の身体的特徴、及び母体に関するデータを用いる。データは省略する。(Marcello,Kimberlee[4]) このデータは共変量が新生児の頭囲、新生児の身長、妊娠期間、母体年齢の 4 つであり、変量は新生児の体重である。群数は妊娠中毒症か否かの 2 群である。このデータを用いて、次の 7 つの方法で変数を選択し、それぞれの自由度調整済み寄与率及び各検定の P 値の比較をする。(井上・桑山 [1], 木村・山原 [3])

1. 変数選択を行わない
2. 各共変量における t 検定
3. 変数増加法
4. 変数減少法
5. 変数増減法
6. 変数減増法
7. トレランス (0.05 以下) を用いて選択

表 1: 各選択法における自由度調整済み寄与率の違い

	選択した変数	R_a^2
方法 1,7	x_1, x_2, x_3, x_4	0.753
方法 2〜6	x_1, x_2	0.756

表 2: 各検定における P 値の違い

	平行性の検定	有意性の検定	共分散分析
方法 1,7	0.012	< 0.0001	0.393
方法 2〜6	0.003	< 0.0001	0.449

表 1, 表 2 から、今回用いたデータでは変数選択によって変数を少なくした場合、平行性の検定と有意性の検定では検出力が高くなり、共分散分析では弱くなると言う結果になった。このデータでは変数が 4 つであり、その全ての影響を取り除くと共分散分析の検出力は高くなると考えられる。一方で、変数選択を行い変数を少なくすると取り除く影響が変数選択を行わない場合よりも小さくなることから、検定結果が保守的になると言える。しかし、自由度調整済み寄与率は高くなることから、変数選択を行わずに全ての変数の影響を取り除くことは、本来存在していないはずの差を見出してしまう可能性があり、変数選択を行う方がよい結果を得ることができると考えられる。

6 プログラム

共分散分析は S-plus 及び R には登録されていないことから、本研究では共分散分析のプログラムを作成した。プログラムでは、変数選択を用いる可能性を考慮して、各共変量の t 検定の有無と各検定の P 値を自由に設定できるようにした。

7 おわりに

本研究では共変量の重要性と共分散分析を用いることの必要性を示唆している。薬効データの解析などでは、人間のデータを扱うため、各々の年齢、性別などがデータに影響を与えることがしばしばある。そのため、共分散分析を用いることで、通常の検定を行うよりも、より精度の高い結果を得ることができると言える。また、幾つかの変数選択の方法を用いて、その精度を比較した結果、変数選択を行うことによってわずかながら分析の精度を上げることができるが、平行性の検定及び、有意性の検定では検出力が高くなり、共分散分析を適応することが難しくなることが明らかになった。一方で、共分散分析の検出力は低くなり、結果が保守的になる可能性があることが明らかになった。

参考文献

[1] 井上勤, 桑山智裕,S-plus における回帰分析の変数選択関数の作成,「南山大学経営学部情報管理学科卒業論文要旨集」,2001.

[2] 奥野忠一, 久米均, 芳賀敏郎, 吉沢正, 「多変量解析法 (改訂版)」, 日科技連出版社,1981.

[3] 木村学, 山原強志,S-plus における回帰分析の変数選択の研究,「南山大学経営学部情報管理学科卒業論文要旨集」,2000.

[4] Marcello Pagano,Kimberlee Gauvreau 著 竹内正弘監訳, ハーバード大学講義テキスト生物統計学入門, 丸善,2003.

[5] P.Armitage,G.Berry 著 椿美智子, 椿広計共訳, 「医学研究のための統計的方法」, サイエンティスト社,2001.