多変量ゼミ §1.11回帰係数の推定と検定(重回帰の場合)

髙見澤 真央

　ここでは，単回帰の流れをもとに，重回帰における回帰係数や定数項の確率分布，母集団での,の信頼区間，検定について考える．

　おおまかな流れとして，

　(1)重回帰モデルの回帰係数と定数項の推定値から，期待値や分散を計算する

　(2) (1)で得られた結果から，信頼区間を求め，検定を行う

　(3)回帰の有意性の検定問題を考える

　このように進めていく．

１.回帰係数の推定と検定(重回帰の場合)

1.1. 回帰係数および定数項の推定値とその期待値，分散

　説明変数の数を として，母集団において次のような回帰モデルが成り立つと仮定する．

ここで，は調査により指定される指定変数，は独立な確率変数で，正規分布に従うものとする．標本のデータ (は標本の個数を表す)を得たとき，これにもとづく回帰係数 は，次のように表される．(§1.5線形重回帰の範囲より)

列

式(2)の分母は，分散共分散行列の行列式で，分子は，分母の行列式の列を で置き換えたもの．

|  |
| --- |
| の分散共分散行列を次のように定義できる  ただし， |

式(2)の分子の行列式に対し列目に関して余因子展開を行い，式(2)を変換する．

|  |
| --- |
| 余因子展開とは・・・  　とする．次正方行列の第 行と第 列を取り除いてできる次正方行列の行列式を倍した数    これを，の の余因子といい，で表す．  さらに，次正方行列に対して，行列式は次のような展開が可能である．  　余因子展開    (第 行に関する展開)  この変換を余因子展開という． |

　(3)の分散共分散行列の行列式を，その 行と 列の余因子をとする．は次のようになる．

式(2)の分子の行列式は，第 列に関して次のように余因子展開できる．

列

よって，式(2)は次のように表せる．

　ここで，の逆行列の要素をと表せば，逆転公式より，次の関係が成り立つ．

|  |
| --- |
| 逆転公式とは・・・  次正方行列におけるの余因子を成分にもつ次正方行列の転置行列を，の余因子行列といい，で表す．すなわち，  　このとき，が正則ならば，次の関係が成り立つ．  　 逆転公式  　これを，逆転公式という．は，正方行列の逆行列．  なお，両辺の行列の要素ごとの関係は次のようになる．  ここで，は逆行列の成分である．余因子行列を求める過程で，行列を転置させるため，対応する要素の行と列が反転する． |

　(5)の関係から，式(4)は次のように変換できる．

|  |  |  |
| --- | --- | --- |
|  |  | |
|  |  | |
|  |  | |
|  |  | |
|  |  | |
|  |  |  |

また，定数項は，次のように変換できる．

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | (§1.5線形重回帰の範囲より) | | |
|  |  | | | |
|  |  | | (式(6)より) | |
|  |  | | | (7) |

　したがって， は正規分布に従う変数, の1次式で表されることがわかる．

これより，回帰係数および定数項の期待値と分散，共分散を求めると次のようになる．

|  |  |  |
| --- | --- | --- |
|  |  | (8) |
|  |  | (9) |
|  |  | (10) |
|  |  | (11) |
|  |  | (12) |
|  |  | (13) |

　式(8)，式(11)より，は母集団における値に対して不偏推定値になっていることがいえる．

　期待値と分散，共分散を求める過程は，長くなるため2.1.にまとめて示す．

1.2. 回帰係数および定数項の信頼区間と検定

　単回帰のときと同様に

|  |  |
| --- | --- |
|  |  |

このように標準化することで，は標準正規分布に従う．

このとき，分母に含まれる未知の誤差分散を不偏推定値

で置き換えて得られる統計量は，1.1.で求めた期待値と分散(式(8),(9),(11),(12))を用いて，

|  |  |  |
| --- | --- | --- |
|  |  | (15) |
|  |  | (16) |

となる．これらは，いずれも自由度の分布に従う．

　式(14)の­が，誤差分散の不偏推定値であることの証明は，2.3.で行う．

　また，式(14)右辺の分子のは，次のような変換できる．

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | (§1.5線形重回帰の範囲より) | |
|  |  | | (17) |

※この変換によって，式(17)の簡潔な表現になるが，導出まではそこそこ複雑．導出過程は2.2.に示す．

　ここから，の信頼区間および仮説あるいは の検定(は与えられた値)は次のようになる．

　信頼区間は，平均して100回中100()回 母集団の を含むことが保証された範囲のこと．

●  **の信頼率の信頼区間：**

● **の信頼率の信頼区間：**

● **仮説の検定()：**

　ならば，危険率で仮説を棄却し，不等号の向きが逆ならば，仮説を採択する．

● **仮説の検定：**

　ならば，危険率で仮説を棄却し，不等号の向きが逆ならば，仮説を採択する．

1.3. 回帰の有意性の検定

　ここでは，回帰の有意性，言いかえると，とりあげた説明変数が全体としての予測に役立つのかについて検定を考える．そのため，観測値の変動(平方和)を回帰による変動とそれ以外に分解していく．

　まず，予測値 は指定変数の組に対して，

このように計算される．

観測値の変動(平方和)は，

で表され，次のように分解される．

|  |  |  |
| --- | --- | --- |
|  |  | |
|  | | |
|  |  | |
|  |  | (23) |

　式(23)の第3項について，

|  |  |  |
| --- | --- | --- |
|  |  | |
|  |  | |
|  |  | (24) |
|  |  |
|  |  |
|  |  |
|  |  |

§1.5線形重回帰の範囲で，次の関係が成り立つことがわかっている．

|  |  |
| --- | --- |
|  | (25) |
|  |
|  |
|  |
|  |
|  |

これによって，式(24)は，

|  |  |  |
| --- | --- | --- |
|  |  | (26) |

これより，式(23)は，

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | |  | | | | (27) |
|  | 全変動() |  | 回帰による変動() | 回帰からの  残差変動() |  | |

このとき，左辺と全変動()する．右辺の第1項は回帰式にもとづく予測値の変動，第2項は残差(残差=観測値予測値)の変動である．前者は，全変動うち回帰によって説明される部分で，後者は説明されない部分にあたる．

とりあげた説明変数が の予測に有効であるとすれば，が大きく(全変動()は説明変数のとりかたに依存せず一定であるから，は小さく)，逆に無効であるとすれば，は小さく(は大きく)なると期待される．

　ここで，

とおけば，は全体の変動のうち回帰によって説明される部分の大きさの割合を表し，その意味から　　決定係数(coefficient of determination)あるいは寄与率と呼ばれる．また，は最小化された予測誤差の平方和に等しい．

§1.7重相関係数の範囲より，重相関係数は，次のように表せる．

　重相関係数を2乗すると，

|  |  |  |
| --- | --- | --- |
|  |  | (29) |

このとき，

|  |  |  |
| --- | --- | --- |
|  |  | |
|  |  | |
|  |  | |
|  |  | (30) |

式(30)の第1項について

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | |  |
|  |  | | |
|  |  | | |
|  |  | (式(25)より) | |

よって，式(30)は，

|  |  |  |
| --- | --- | --- |
|  |  | |
|  |  | (31) |

このような関係が成り立つ．

式(31)を用いて，式(29)を変換すると，次のようになる．

|  |  |  |
| --- | --- | --- |
|  |  | (29)再掲 |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  | (32) |

　式(30)より，寄与率は，重相関係数の2乗に等しいことがわかる．

　モデルが適合しているとき，は自由度のカイ2乗分布に従う．また，とくに説明変数が の予測に寄与しない，すなわち母集団における回帰係数の値が のときには，もとは独立に自由度のカイ2乗分布に従うことが知られている．これより，次のような分散分析表が構成される．

**分散分析表**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 変動要因 | 平方和 | 自由度 | 不偏分散 | 分散比 |
| 回帰による |  |  |  |  |
| 回帰からの |  |  |  |  |
| 全体 |  |  |  |  |

が普通より大きい，つまり回帰による変動が残差の変動を凌駕していれば，重回帰式は無意味ではない．

この分散分析表において，分散比が

ならば，仮説は危険率で棄却され，とりあげた説明変数は全体としての予測に役立つと結論付けられる．このとき，回帰は有意であるという．

ここで，は自由度の分布の上限点である．複数の係数についての仮説を検定したいときに，検定がよく用いられる．

|  |
| --- |
| 分布とは・・・  　2つの確率分布が次の条件を満たすとする．  は自由度のカイ2乗分布に従う  は自由度のカイ2乗分布に従う  は独立である．  　ここで，とをそれぞれの自由度で割って調整した後にとった比，すなわちフィッシャー分散比を  と定義すると，が従う確率分布を自由度の分布といい，またはと表す． |

2. 式変換の補足

2.0. 式変換に用いるδ関数および統計量

　式変換を行うにあたり，δ関数を定義し，そこから導ける様々な関係を示す．

　これを，δ関数として定義する．

　ここで， と と 次単位行列について

すなわち

行番号と列番号が一致するときにのみ１をとることから，

が成り立つ．また，δ関数は，定数について，

が成り立つ．

　続いて，に関する統計量を計算していく．期待値は，

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | (式(1)より) | |
|  |  | (なお ) | |
|  |  | | (37) |

の共分散は，

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | | |
|  |  | | |
|  | (式(37)より) | | |
|  |  | | |
|  |  | (より) | |
|  |  | | (38) |

について，は互いに独立で無相関であるため，

定数 とすると，

が成り立つ．共分散 について，

が成り立つ．

これらの式を用いて，変換を行っていく

2.1. 期待値と分散，共分散の計算

式(8) の導出

|  |  |
| --- | --- |
|  |  |
|  |  |
|  |  |
|  | (式(37) より) |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

　式(10) の導出

|  |  |
| --- | --- |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  | (式(38) より) |
|  |  |
|  | ( |
|  |  |
|  |  |
|  |  |
|  |  |

　式(9) の導出

　式(10)において が のとき，式(9)となるため，

|  |  |
| --- | --- |
|  |  |
|  |  |

　式(10) の導出

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  |  | |
|  |  | |
|  |  | |
|  |  | |
|  |  | |
|  | (式(37) より) | |
|  |  | |
|  |  | |
|  |  | |
|  |  | |

　式(12) の導出

|  |  |  |
| --- | --- | --- |
|  |  | |
|  |  | |
|  | () | |
|  |  | |
|  |  | |
|  |  | |
|  |  | |
|  |  | |
|  |  | |
|  |  | |
|  |  | |
|  |  | (41) |

　式(41)の第1項について

|  |  |  |
| --- | --- | --- |
|  |  | |
|  |  | |
|  | (式(1) より) | |
|  |  | |
|  |  | |
|  |  | |
|  |  | |
|  |  | (42) |

また，式(41)の第2項について

|  |  |  |
| --- | --- | --- |
|  |  | |
|  |  | |
|  |  | |
|  | (式(1) より)  () | |
|  |  | |
|  |  | |
|  |  | |
|  | (式(38) より) | |
|  |  | |
|  |  | |
|  |  | |
|  |  | (43) |

　式(41)に式(42)と式(43)を代入して

|  |  |
| --- | --- |
|  |  |
|  |  |
|  |  |

　式(13) の導出

|  |  |
| --- | --- |
|  |  |
|  |  |
|  | () |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  | (式(43) )　 (式(10) ) |

2.2. 式(17) の証明

|  |  |  |
| --- | --- | --- |
|  |  | |
|  |  | |
|  | () | |
|  |  | |
|  |  | (44) |

このとき，式(44)の第1項は，

式(44)の第2項は，

|  |  |  |
| --- | --- | --- |
|  |  | |
|  |  | (46) |

式(44)の第3項は，

|  |  |  |
| --- | --- | --- |
|  |  | |
|  |  | |
|  |  | |
|  | () | |
|  |  | |
|  |  | |
|  | () | |
|  |  | |
|  | () | |
|  |  | |
|  |  | |
|  |  | (47) |

　よって，式(44)に，式(45)，式(46)，式(47)を代入して，

|  |  |
| --- | --- |
|  |  |
|  |  |

以上で，式(17)が示された．

2.3. が誤差分散の不偏推定値であることの証明

　式(14) が誤差分散の不偏推定値であることを示すため，の期待値を求める．

|  |  |  |
| --- | --- | --- |
|  |  | |
|  |  | |
|  | (式(1) より) | |
|  |  | |
|  |  | |
|  |  | |
|  |  | |
|  |  | |
|  |  | |
|  |  | |
|  |  | |
|  | ()　 () | |
|  |  | (48) |
|  |  |

このとき，

|  |  |  |
| --- | --- | --- |
|  |  | |
|  |  | |
|  |  | |
|  | (式(7) より) | |
|  |  | |
|  |  | |
|  |  | |
|  | (式(37) より) | |
|  |  | |
|  |  | (49) |
|  | () | |

　式(49)より，式(48)の第4項は，

|  |  |  |
| --- | --- | --- |
|  |  | |
|  |  | |
|  |  | |
|  |  | (50) |

式(49)と同様に，

|  |  |  |
| --- | --- | --- |
|  |  | |
|  |  | (51) |

ここでの導出過程は，式(49)の繰り返しになる部分が多いため，省略する．

式(51)より，式(48)の第5項は，

|  |  |  |
| --- | --- | --- |
|  |  | |
|  |  | |
|  |  | |
|  |  | |
|  |  | |
|  |  | (52) |

式(48)の第1項は，

|  |  |  |
| --- | --- | --- |
|  |  | |
|  | (式(14) より) | |
|  |  | (53) |

式(48)の第2項は，

|  |  |  |
| --- | --- | --- |
|  |  | |
|  | (式(14) より) | |
|  |  | |
|  |  | (54) |

式(48)の第3項は，

|  |  |  |
| --- | --- | --- |
|  |  | (55) |

式(48)の第6項は，

|  |  |  |
| --- | --- | --- |
|  |  | |
|  |  | |
|  |  | (56) |

最後に，式(48)に式(50)，式(52)，式(53)，式(54)，式(55)，式(56)を代入して，

|  |  |  |
| --- | --- | --- |
|  |  | |
|  |  | |
|  |  | |
|  |  | |
|  |  | |
|  |  | |
|  |  | |
|  |  | |
|  |  | (57) |

したがって，

これより，

が誤差分散の不偏推定値であることがわかる．