# 目　次

率：第２章参照）で棄却できることを示している。要するに，すべての回帰係数が０になる確率はきわめて小さいということである。しかし，このことは，得られた回帰式で観測値を危険率5%未満で予測できるということを意味しているわけではない。ここでの有意確率は，危険率5%未満で帰無仮説を棄却でき，「すべての回帰係数は０ではない」という対立仮説を受容してよいことを示しているだけである。このところを間違えないようにしよう。得られた重回帰式が有効であるかどうかは，回帰係数の検定を行ってからでないと判断できない。そこで，次に以下の出力表（**図表3-29**）に目を向けることにしよう。

**係数**[a]

| モデル | | 非標準化係数 | | 標準化係数 | | | Bの95%信頼区間 | | 相関係数 | | | 共線性の統計量 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | B | 標準誤差 | ベータ | t | 有意確率 | 下限 | 上限 | ゼロ次 | 偏 | 部分 | 許容度 | VIF |
| 1 | (定数) | 13.007 | 3.586 | | 3.627 | .000 | 5.943 | 20.072 | | | | | |
| | 年齢 | -.005 | .049 | -.007 | -.106 | .916 | -.103 | .092 | .017 | -.007 | -.007 | .968 | 1.033 |
| | 世帯人数 | -.045 | .184 | -.016 | -.243 | .809 | -.408 | .318 | .002 | -.016 | -.015 | .977 | 1.024 |
| | 年間世帯所得 | .002 | .001 | .197 | 3.081 | .002 | .001 | .003 | .195 | .195 | .195 | .977 | 1.023 |

a 従属変数: 老後幸福感

$t = \dfrac{B}{\text{標準誤差}}$

従属変数（観測値）との相関関係

**相関係数**[a]

| モデル | | | 年間世帯所得 | 世帯人数 | 年齢 |
|---|---|---|---|---|---|
| 1 | 相関係数 | 年間世帯所得 | 1.000 | -.076 | -.120 |
| | | 世帯人数 | -.076 | 1.000 | -.122 |
| | | 年齢 | -.120 | -.122 | 1.000 |
| | 共分散 | 年間世帯所得 | .000 | .000 | .000 |
| | | 世帯人数 | .000 | .034 | -.001 |
| | | 年齢 | .000 | -.001 | .002 |

a. 従属変数: 老後幸福感

回帰係数の相関行列

回帰係数の分散－共分散行列

分散

**共線性の診断**[a]

| モデル | 次元 | 固有値 | 条件指標 | 分散の比率 | | | |
|---|---|---|---|---|---|---|---|
| | | | | (定数) | 年齢 | 世帯人数 | 年間世帯所得 |
| 1 | 1 | 3.515 | 1.000 | .00 | .00 | .02 | .02 |
| | 2 | .316 | 3.333 | .00 | .00 | .13 | .90 |
| | 3 | .165 | 4.610 | .01 | .01 | .85 | .07 |
| | 4 | .003 | 33.605 | .99 | .99 | .00 | .01 |

a. 従属変数: 老後幸福感

$= \sqrt{\dfrac{3.515}{0.316}}$

図表3－㉙

「**係数**」の表にある「**非標準化係数**」(B) が得られた重回帰式の「定数項」(切片) と「回帰係数」である。重回帰分析の場合の回帰係数を一般に偏回帰係数と呼ぶ。これを一次関数の形で表現すると次のようになる。

老後幸福感得点
$=13.007+(-0.005)\times$年齢$+(-0.045)\times$世帯人数$+0.002\times$年間所得
$=13.007-0.005\times$年齢$-0.045\times$世帯人数$+0.002\times$年間所得

この重回帰式が意味するところは，老後幸福感得点は，年齢が１歳上がると0.005倍低下し，世帯人数が１人増加すると0.045倍低下し，年間所得が１万円増加すると0.002倍増加する，ということである。

例えば，年齢が70歳で，世帯人数が２人，年間所得が400万円であれば，老後幸福感の予測得点は以下のようになる。

$$A = -97.052 + 0.666 \times \text{年齢} + 6.061 \times \text{居住年数} + 3.466 \times \text{趣味の数} + 0.367 \times \text{通勤時間}$$

$$B = -74.558 + 0.439 \times \text{年齢} + 4.900 \times \text{居住年数} + 5.537 \times \text{趣味の数} + 0.432 \times \text{通勤時間}$$

これらの式に，先ほどの「年齢が58歳で居住年数が10年，趣味が４つあり，通勤時間は50分」の人のデータを入れると，$A = 34.973$，$B = 42.057$ となる。$B$ の値の方が大きいので，この人はB群に分類される。このように，分類関数を用いるときには，算出された値の大きい方に分類する。3群以上の場合は最大値のグループに分類する。

**図表4-21** の「**ケースごとの集計**」には，各ケースについて，「実際のグルー

ケースごとの統計

| | ケース番号 | 実際のグループ | 最大グループ 予測グループ | 条件確率 P(D>d \| G=g) p | 自由度 | 事後確率 P(G=g \| D=d) | 重心へのMahalanobisの距離の2乗 | 2番目のグループ グループ | P(G=g \| D=d) | 重心へのMahalanobisの距離の2乗 | 判別得点 関数1 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 元のデータ | 1 | 1 | 1 | .204 | 1 | 1.000 | 1.610 | 2 | .000 | 31.123 | 3.424 |
| | 2 | 1 | 1 | .463 | 1 | .998 | .538 | 2 | .002 | 12.792 | 1.422 |
| | 3 | 1 | 1 | .578 | 1 | .999 | .309 | 2 | .001 | 14.094 | 1.599 |
| | 4 | 1 | 1 | .772 | 1 | 1.000 | .084 | 2 | .000 | 21.156 | 2.445 |
| | 5 | 1 | 1 | .116 | 1 | .925 | 2.468 | 2 | .075 | 7.501 | .584 |
| | 6 | 1 | 1 | .291 | 1 | 1.000 | 1.114 | 2 | .000 | 28.788 | 3.210 |
| | 7 | 1 | 1 | .806 | 1 | 1.000 | .060 | 2 | .000 | 20.752 | 2.400 |
| | 8 | 2 | 2 | .179 | 1 | .970 | 1.808 | 1 | .030 | 8.792 | -.810 |
| | 9 | 2 | 2 | .891 | 1 | 1.000 | .019 | 1 | .000 | 19.776 | -2.292 |
| | 10 | 2 | 2 | .310 | 1 | 1.000 | 1.029 | 1 | .000 | 28.347 | -3.169 |
| | 11 | 2 | 2 | .179 | 1 | 1.000 | 1.803 | 1 | .000 | 31.953 | -3.498 |
| | 12 | 2 | 2 | .943 | 1 | 1.000 | .005 | 1 | .000 | 17.963 | -2.083 |
| | 13 | 2 | 2 | .283 | 1 | .991 | 1.153 | 1 | .009 | 10.473 | -1.081 |
| | 14 | 2 | 2 | .996 | 1 | 1.000 | .000 | 1 | .000 | 18.537 | -2.151 |
| 交差確認済み[a] | 1 | 1 | 1 | .181 | 4 | 1.000 | 6.252 | 2 | .000 | 40.496 | |
| | 2 | 1 | 1 | .382 | 4 | .991 | 4.185 | 2 | .009 | 13.599 | |
| | 3 | 1 | 1 | .022 | 4 | .966 | 11.402 | 2 | .034 | 18.079 | |
| | 4 | 1 | 1 | .971 | 4 | 1.000 | .523 | 2 | .000 | 19.938 | |
| | 5 | 1 | 1 | .360 | 4 | .914 | 4.357 | 2 | .086 | 9.073 | |
| | 6 | 1 | 1 | .443 | 4 | 1.000 | 3.735 | 2 | .000 | 32.919 | |
| | 7 | 1 | 1 | .360 | 4 | 1.000 | 4.354 | 2 | .000 | 22.958 | |
| | 8 | 2 | 2 | .167 | 4 | .871 | 6.473 | 1 | .129 | 10.301 | |
| | 9 | 2 | 2 | .001 | 4 | .999 | 19.625 | 1 | .001 | 34.266 | |
| | 10 | 2 | 2 | .305 | 4 | 1.000 | 4.830 | 1 | .000 | 33.951 | |
| | 11 | 2 | 2 | .026 | 4 | 1.000 | 11.027 | 1 | .000 | 50.406 | |
| | 12 | 2 | 2 | .520 | 4 | 1.000 | 3.229 | 1 | .000 | 18.702 | |
| | 13 | 2 | 1** | .009 | 4 | .506 | 13.454 | 2 | .494 | 13.505 | |
| | 14 | 2 | 2 | .026 | 4 | .999 | 11.024 | 1 | .001 | 25.062 | |

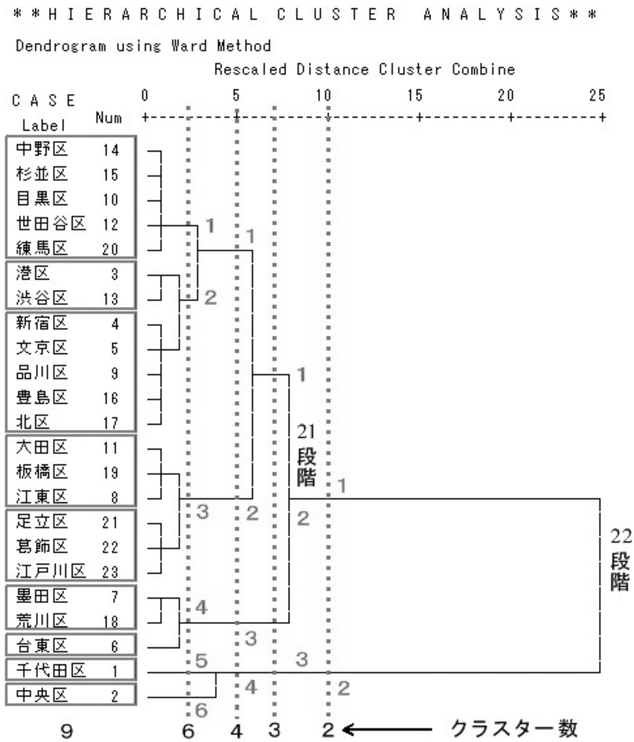
元のデータでは、Mahalanobis の距離の 2 乗は正準関数に基づきます。
交差確認済みのデータでは、Mahalanobis の距離の 2 乗は観測値に基づきます。

**. 間違って分類されたケース

a. 交差確認は分析中のケースのみに実行されます。交差確認では、各ケースはそのケース以外のすべてのケースから得られた関数により分類されます。

図表4-21

第4章 類型化と分類 ― 判別分析とクラスター分析 ―




図表4-39

らである。したがって，クラスター数を検討する際には，そうした結合が生じる前の段階までのクラスター化の過程に目を向ければよいことになる。

以上のようなクラスター化の過程は「**デンドログラム**」として表現される（**図表4-39**）。デンドロ（dendro）というのはギリシャ語で樹木を意味する言葉で，デンドログラム（dendrogram）は樹状図とも呼ばれる。これがなんで樹木の形をした図なんだ，と思う人は，この図を右に90度回転させてみよう。上に並ぶ個々のケースが葉っぱに，クラスター化の過程を示す線で結ばれた部分が下へ行くほど太くなっていく枝に，そして，最終的に全部のケースが結合された部分は太い幹のように見えないだろうか。そう見えた人が（誰だか知らないが），こうした図をデンドログラムと呼ぶようになったのが，この言葉の始まりである。

「**クラスタ凝集経過工程**」の第１段階で結合された 14（中野区）と 15（杉並

になってしまう。非階層的クラスター分析が採用されるのは，一つにはそうした計算上の制約からである。そこで，計算不能にならないケース数までは階層的クラスター分析を実行すればよいと言うこともできる。現在では中程度のパソコンでもかなりのケースの階層的クラスター分析が可能と考えられるので，非階層的クラスター分析については簡単に触れておくだけにする。

非階層的クラスター分析の代表的なものは K-means 法（K-平均法）である。SPSS でいう「大規模ファイルのクラスタ」分析もこの方法を採用している。K-means 法の K というのはあらかじめ指定したクラスターの数のことである。K 個のクラスターに分類するということを先に決めて，暫定的に構成されたクラスターの中心とケースとの距離を計算してそれぞれのケースを最も近いクラスターに所属させる。このとき，クラスターの中心として，そのクラスターに所属させたすべてのケースと暫定的な中心（重心）との距離の平均値（mean）を採用する。この方法が K-means と呼ばれるのはそのためである。一度すべてのケースを K 個の各クラスターに所属させたら，再度クラスターの中心を計算してケースの所属を見直す。この操作をすべてのケースについて繰り返し行い，各ケースと所属クラスタの中心との距離の合計が最小になったところで所属クラスタを確定する。距離の測り方には前にも述べたように様々なものがあるが，SPSS の「大規模ファイルのクラスタ」分析ではユークリッド距離を使っている。非階層的クラスター分析は，何らかの根拠や必要からクラスター数があらかじめ決められていて，ケース数も膨大で，クラスターを階層的にとらえる必要がないような場合に威力を発揮する方法といえる。

SPSS には，非階層的クラスター分析の方法としてもう一つ「**TwoStep クラスタ**」分析というのを用意している。これも大規模ファイルのクラスター分析に適した方法であるが，特徴的なことは，変数のタイプを問わないことである。間隔データや比率データのような連続データからなる変数，順序データの変数，名義データのようなカテゴリ変数も使える。そして，クラスター数を自動的に決定するところや作図の出力が豊富なところも「階層クラスタ」や「大規模ファイルのクラスタ」とは異なるところである。「大規模ファイルのクラスタ」や「TwoStep クラスタ」は，ケース数が多くないからといって使えないわけではないから，機会を見て一度試してみたらよい。

合には有意確率が 0.05 未満になることが多く，帰無仮説が棄却されてしまう。そこで，大規模標本の場合には，この方法による適合度検定は有用ではないとされている。そのような時には，RMSEA（Root Mean Square Error of Approximation：平均 2 乗誤差平方根）という検定方法を試してみるとよい。RMSEA は，出力された「**適合度検定**」の表にあるカイ 2 乗値と自由度を次の式に代入すれば得られる。

$$\mathrm{RMSEA}=\sqrt{\left(\frac{\text{カイ2乗値}}{\text{自由度}}-1\right)\div(\text{サンプル数}-1)}$$

一般に，RMSEA が 0.08 未満であれば適合度は良好で，0.05 未満なら文句なしと判断される。なお，カイ 2 乗値が自由度よりも小さいときには RMSEA を 0 とみなす。RMSEA は標本規模に制約されないから小規模標本でも利用できる。ここでの例で言えば，カイ 2 乗値の 24.180 は自由度の25よりも小さいから RMSEA は 0 ということになる。SPSS でもカイ 2 乗値による適合度検定だけでなく RMSEA を出力してくれればよいのにと思う。

**図表5-39** の「**パターン行列**」（因子パターン行列）は，ここで採用した直接オブリミン法やプロマックス回転のような斜交回転の場合に出力される回転

**▼図表5−㊴**　パターン行列 a

| | 因子 | | |
|---|---|---|---|
| | 1 | 2 | 3 |
| 私はクラブやグループ、団体のためにがんばっている | .946 | | |
| 私は地域のためにがんばっている | .941 | | |
| 私は社会や世の中のためにがんばっている | .874 | | |
| 私を正しく評価できない人は、ろくな人物ではない | | .878 | |
| 私は自分のことをなかなかの人間だと思っている | | .873 | |
| 私は人と議論しても負けることが少ない | | .790 | |
| 私にまかせておけば、うまくいくことが多い | | .723 | |
| 私には良いところがたくさんある | | .610 | |
| 私は子どものためにがんばっている | | | .846 |
| 私は友人に信頼されている | | | .800 |
| 私は家族の者に頼りにされている | | | .474 |

因子抽出法：最尤法
回転法：Kaiser の正規化を伴うオブリミン法
a. 4 回の反復で回転が収束しました。

## 6-5-4 一元配置分散分析の出力

最初に出力されるのは「**記述統計**」である（**図表6-32**）。「グループの平均」を実行したときに出力された「報告書」にはなかった各グループの「平均値の信頼区間」が表示されている。

図表6-32

現在の給与　　記述統計

| | 度数 | 平均値 | 標準偏差 | 標準誤差 | 平均値の 95% 信頼区間 下限 | 平均値の 95% 信頼区間 上限 | 最小値 | 最大値 |
|---|---|---|---|---|---|---|---|---|
| 1　25歳未満 | 83 | $25,314.94 | $4,224.578 | $463.708 | $24,392.48 | $26,237.40 | $18,150 | $40,800 |
| 2　25－29 | 154 | $38,476.62 | $15,244.107 | $1,228.405 | $36,049.80 | $40,903.45 | $21,900 | $91,250 |
| 3　30－39 | 92 | $45,084.95 | $21,600.983 | $2,252.058 | $40,611.51 | $49,558.38 | $20,850 | $110,625 |
| 4　40－49 | 68 | $32,920.88 | $20,023.380 | $2,428.191 | $28,074.19 | $37,767.57 | $16,950 | $135,000 |
| 5　50歳以上 | 76 | $24,565.13 | $7,448.973 | $854.456 | $22,862.97 | $26,267.29 | $15,750 | $65,000 |
| 合計 | 473 | $34,418.45 | $17,093.723 | $785.970 | $32,874.01 | $35,962.88 | $15,750 | $135,000 |

図表6-33

現在の給与　　等分散性の検定

| Levene 統計量 | 自由度1 | 自由度2 | 有意確率 |
|---|---|---|---|
| 28.216 | 4 | 468 | .000 |

次に出力されるのは「**等分散性の検定**」結果である（**図表6-33**）。「**有意確率**」が 0.000 と小さく，「各グループの分散は等しい」という帰無仮説は棄却される。したがって，このデータを用いて分散分析を実行することは適切ではないことになる。そこで，分散分析によらずに，「**平均値同等性の耐久検定**」という恐ろしげな表題の付いた「**Brown-Forsythe**」と「**Welch**」の検定結果の表からグループ間に有意差があるかないかを判断することにする（**図表6-34**）。表の下に「漸近的 F 分布」とあるのは，サンプルサイズ（ケース数）が大きくなると，それらの統計量は F 分布に近づきますよ，という注記である。「**有意**」の欄を見ると，いずれの有意確率も 0.000 と小さいので，グループによって平均値に有意な差があると判断してよいことになる。

図表6-34

現在の給与　　平均値同等性の耐久検定

| | 統計 | 自由度1 | 自由度2 | 有意 |
|---|---|---|---|---|
| Welch | 44.282 | 4 | 199.842 | .000 |
| Brown-Forsythe | 29.251 | 4 | 250.687 | .000 |

a. 漸近的 F 分布

「**分散分析**」の表に示されている 1 次の項の「重み付き」は，前節で実行した「グループの平均」で出力された「分散分析表」の中の「線型成分」と同