

統計的仮説検定 その3

1 多重比較

前節で学習した分散分析は、3つ以上の平均値の中に異なる母集団からの標本がある、すなわち有意に異なっているものがあることを検出することができる。しかし、どの平均値が他とは異なっているのかについては、分散分析により求めた F の値からだけではわからない。そこで、それぞれの平均値間の差が有意であるかどうかをすべての組み合わせについて行うことを**多重比較**という。これまでに多くの方法が考案されており、標本の分布や等分散性などの違いによりどれを使用すべきかについては大体決まっている。この授業では、等分散性を前提に以下の2つの方法について学習する。

1.1 区間推定を利用する方法

1.1.1 区間推定とは

標本から母平均を推定するのに最も確からしいのは標本平均を求めることであることはすでに学習した。しかし、標本平均は標本を採るたびに変動するので、測定ごとに異なるのが普通である。そのため、標本数を大きく取れば母平均の推定値としての確からしさが増大するが、様々な制約から無限に標本を集めることは難しい。そこで、測定値から母平均が存在する区間を推定することで、ある確率での母平均の値の範囲を推定することが可能となる。

母標準偏差がわかっていない場合、標本から得られる母数の推定値は標本平均と不偏分散（もしくはその平方根である標準偏差）である。また、標本の値は t 分布に従うことも予想される。そこで、これらの値から、例えば、95%の確率で母平均が含まれると予想される範囲を求めることができる。例えば、以下のようなデータがあるとする。

2.15 2.20 2.18 2.21 2.09 2.24 2.09

これら7つの値の平均値 \bar{x} は 2.166 であり、不偏分散 v^2 は 0.003429、標準偏差 v は 0.05855 である。このとき、 t 分布で両端の 2.5% ずつに含まれる値の境界値を求めてみる。今、自由度 ϕ は 6 であるので、それに該当する t の値を見てみると、表1のようになっている。

表 1: 自由度 $\phi = 6$ のときの α に対する t の値

α	0.05	0.02	0.01	0.001
t	2.4469	3.1427	3.7074	5.9588

よって、 t 分布の定義式 $t = (\bar{x} - \mu)/(v/\sqrt{n})$ から、平均値からのずれ量は以下の式となる。

$$2.4469 \times \frac{0.05855}{\sqrt{7}} = 0.05415 \quad (1)$$

これより、95%の確率で平均値が含まれると推定される範囲は 2.166 ± 0.054 となるので、2.112 から 2.220 の範囲となる。一般的には、

$$\bar{x} - t(\phi, \alpha) \frac{v}{\sqrt{n}} \leq \mu \leq \bar{x} + t(\phi, \alpha) \frac{v}{\sqrt{n}} \quad (2)$$

の範囲内に母平均 μ があると推定されるので、その他の α についても求めると、表2のような範囲と推定される。区間推定の精度を上げる、すなわち α 値を小さくすると、当然ながら推定範囲は広くなるのがわかる。

表 2: サンプルデータの母平均の区間推定

α	0.05	0.02	0.01	0.001
ずれ量	0.054	0.070	0.082	0.132
下限	2.112	2.096	2.084	2.034
上限	2.220	2.235	2.248	2.298

1.1.2 差の有意性の確認

前項で見たように、平均値と不偏分散から母平均の区間推定ができるということは、分散分析において使用した平均値についても区間推定を行うことにより、それぞれの平均値の推定範囲を求めることが可能であるということである。そうすると、推定区間が重ならない平均値については、 $1 - \alpha$ の確率で母平均が異なっていることが言える。すなわち、その平均値間の差が有意であると言うことができる。ただし、分散分析においては、自由度の考え方が単純な区間推定の際と異なり $N - k$ を使用すること、そして、分散としては郡内変動の分散 V_W を使用するところがポイントである。よって、式 (2) は以下のように書き換えられる。

$$\bar{x}_i. - t(N - k, \alpha) \sqrt{\frac{V_W}{n_i}} \leq \mu_i \leq \bar{x}_i. + t(N - k, \alpha) \sqrt{\frac{V_W}{n_i}} \quad (3)$$

以下、前回の演習で使用した値について区間推定により差の有意性について確認してみよう。4つの水準 A1~A4 に関して式 (3) を用いて区間推定を行う。危険率 5% で有意に異なることが分散分析により示されていたので、ここでは $\alpha = .05$ のみについて行ってみる。また、 $V_W = 4.002$ であり、 t の値は、 $t(16, 0.05) = 2.1199$ を使用する。計算により求められた区間を表 3 に示す。

表 3: 先週のデータの区間推定

水準	A1	A2	A3	A4
平均値	13.6	15.2	18.0	14.0
下限	11.70	13.44	15.88	12.10
上限	15.50	16.90	20.12	15.90

表から数値の重なりの有無を見るのはわかりづらいかもしれないが、A1 と A3 の群の間に重なりが無い。これを図に示すと、図 1 のようになるので、少し見やすくなったであろう。つまり、前回例題として試した分散分析のデータにおいては、4つの群のうち、平均値の最も小さい A1 と最も大きい A3 の間に有意な差があったということになる。A4 と A3 の重なりはごくわずかであるが、この方法では有意な差とは言えないことになる。

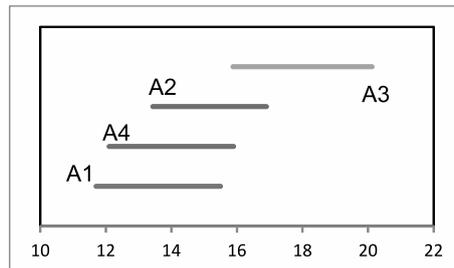


図 1: 各群の区間推定範囲の重なり

なお、このような平均値の比較を行う際、図 2 のように標準誤差 (RMS 誤差) をつけ、さらに有意な差が見いだされた対のところに記号を付与する形式がよく用いられる。アスタリスク (*) の記号は有意差がある場合に使用し、通常 1 個の場合には $p < .05$ を表し、2 個で $p < .01$ 、3 個では $p < .001$ を意味している。

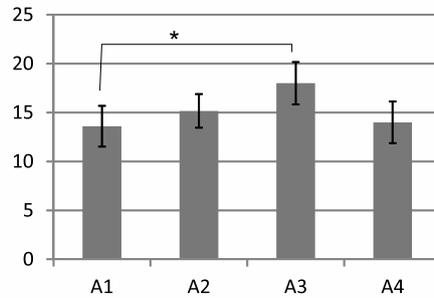


図 2: エラーバー付きのグラフの例

1.2 テューキー法

1.2.1 スチューデント化された範囲の表

前述した区間推定を用いる多重比較はそれぞれの平均値について区間を推定する必要があった。そのため、作業としては少し煩雑になってしまう。そこで、一度に全部の差の有意性を検定する方法がいくつか考案されているが、ここでは代表的な方法のテューキー法について紹介する。Tukey という 19 世紀にアメリカで活躍した数学者の名前を取った方法である。テューキー法では新たにスチューデント化された範囲の表と名付けられた表を用いる。今回配布したものがそれである。有意水準 α 、水準の個数 k 、そして自由度 df によって求められる $q_{\alpha,k,df}$ が表にまとめられている。 q を求める式は複雑で積分なども含まれているため、正規分布表や t 分布表と異なり、表計算ソフトで手軽に作成するわけにはいかない。そこで、この授業でも一般的に公開されているものを使用するので、 α は 0.05 と 0.01 の 2 種類のみである。

1.2.2 臨界差

平均値の間の差がある一定値を超えたらその平均値間に有意な差があると判断する基準として**臨界差**を使用するのがテューキー法である。臨界差 HSD とは、Honestly Significant Difference の略であり、各群の標本数が等しく n であれば以下の式で求められる。

$$HSD = q_{\alpha,k,df} \sqrt{\frac{V_W}{n}} \quad (4)$$

しかし、標本数が各群で異なる場合には以下の式をそれぞれの組み合わせに対して計算する。このように各群の標本数が異なる場合のやり方をテューキー-クラマース法と呼ぶことがある。

$$HSD = q_{\alpha,k,df} \sqrt{\frac{V_W \left(\frac{1}{n_i} + \frac{1}{n_j} \right)}{2}} \quad (5)$$

前節で用いた値を再び利用して臨界差を用いた多重比較を行ってみる。まずは、データを平均値の小さい順に並べ直し、それぞれの差を記述した表 4 を作成する。差は大きい方から小さい方を引く正の値のみで良いので、表は右上のみデータを記入する。

表 4: 平均値の差の表

	A1	A4	A2	A3
平均値	13.60	14.00	15.17	18.00
A1	-	0.40	1.57	4.40
A4		-	1.17	4.00
A2			-	2.83
A3				-

次に、ステューデント化された範囲の表を用いて q を求め、臨界差を計算する。まず、 α については 0.05 とし、群の個数 k は 4 である。自由度 df については、 $N - k$ であるので、 $df = 16$ である。表から $q_{0.05,4,16} = 4.05$ と求められるので、臨界差を計算すると以下の表 5 のようになる。

表 5: 各群間の HSD

	A1-A2	A1-A3	A1-A4	A2-A3	A2-A4	A3-A4
HSD	3.47	3.84	3.62	3.70	3.47	3.84

A1 と A4 の間の HSD は 3.62 であるが、実際の平均値の差はそれを超える 4.40 であり、ここに有意な差があることがわかる。さらに、A3 と A4 の間の HSD が 3.84 であり、実際の差が 4.00 であることからここにも有意な差が見られる。チューキー法を用いた多重比較の結果が前節の区間推定を用いたものとは異なっているが、手法により有意差の検定結果が異なることはよく見られることであり、この場合のようにチューキー法の方が有意差をより多く検出することから、検出力が高いといわれている。図 2 をチューキー法の結果で描き直したものを図 3 に示す。

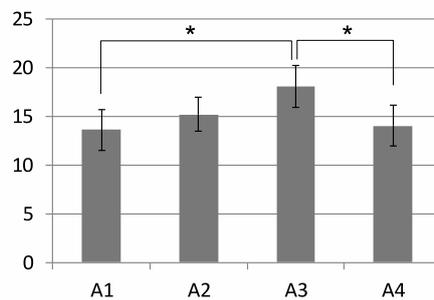


図 3: 図 2 を描き直したもの