Kobe University Repository : Kernel

PDF issue: 2025-05-12

正規母集団の平均の差の検定に関する一考察

難波, 明生

(Citation)
国民経済雑誌,227(4):69-81

(Issue Date)
2023-06-10

(Resource Type)
departmental bulletin paper

(Version)
Version of Record

(JaLCDOI)
https://doi.org/10.24546/0100482441

(URL)
https://hdl.handle.net/20.500.14094/0100482441

ISSN 0387-3129

国民経済雑誌
THE
KOKUMIN-KEIZAI ZASSHI
(JOURNAL OF ECONOMICS & BUSINESS ADMINISTRATION)

正規母集団の平均の差の検定に関する一考察

難　波　明　生

国民経済雑誌　第227巻　第4号　抜刷
2023年6月

神戸大学経済経営学会

『国民経済雑誌』第227巻第4号（2023年6月），69-81

正規母集団の平均の差の検定に関する一考察

難　波　明　生[a]

応用統計学において母集団の平均の差を検定したい状況は多数存在する。しかし，母集団に正規性を仮定した場合でも，平均の差について厳密な有意水準で検定を行うことは容易ではなく，ウェルチの検定と呼ばれる近似的な方法が用いられる。本稿では，ウェルチの検定の有効性をいくつかの側面から検証する。

キーワード　ベーレンス・フィッシャー問題，ウェルチの検定，
シェッフェの方法

1　は　じ　め　に

2つの母集団の平均に差があるかは，応用統計学において頻繁に直面する問題である。例えば，神戸市と芦屋市で平均所得に差があるか，薬の投与をしたグループとしていないグループで差があるかのように，処置群と対象群で差があるかなど，平均の差に興味があるケースは非常に多い。2つの母集団の平均に差があるかを検証するには，平均の差に関する検定を行えば良いのであるが，正規母集団を仮定したとしても，2つの母集団の分散が異なる場合には厳密な有意水準での検定を行うことはできない。このように，2つの母集団の平均（中心）を比較する問題はベーレンス・フィッシャー（Behrens-Fisher）問題と呼ばれる。

正規母集団の平均の差の検定については，標本が十分大きければ，漸近正規性を用いて検定すれば良い。しかし，有限の標本においては正規近似はあまり正確ではない。このため，多くの場合 Welch（1947）により提案されたウェルチの検定という方法を用い近似的な検定が行われる。

本稿ではまず，ウェルチの検定で用いられる統計量の分布関数を導出し，その近似精度をコンピュータを用いた数値計算とシミュレーションにより検証する。次に，数値計算により厳密な有意水準の検定を行う事のできる検定方法と検定力の比較を行う。結果として，ウェルチの検定の近似は多くの場合非常に正確であることと，他の厳密な検定と比較して良好な検定力を持つことが示される。このことから，実際の応用ではウェルチの検定を行うことが

a　神戸大学大学院経済学研究科，namba@kobe-u.ac.jp

非常に有益であると示唆される。

2　ウェルチの検定

2.1　検定法

$X_{1i}(i=1, 2, \ldots, n_1)$ は平均 μ_1，分散 σ_1^2 の正規母集団からの大きさ n_1 の標本，$X_{2i}(i=1, 2, \ldots, n_2)$ は平均 μ_2，分散 σ_2^2 の正規母集団からの大きさ n_2 の標本であるとする。ここで興味があるのは，2つの正規母集団の平均 μ_1 と μ_2 に差があるかどうかである。このことを検証するため，帰無仮説を $H_0: \mu_1=\mu_2$，対立仮説を $H_1: \mu_1\neq\mu_2$ として仮説検定を行うことを考えよう。この時，帰無仮説 H_0 が正しければ

$$\frac{\bar{X}_1-\bar{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} \qquad (1)$$

は標準正規分布 $N(0, 1)$ にしたがう。ただし $\bar{X}_1=\frac{1}{n_1}\sum_{i=1}^{n_1}X_{1i}$，$\bar{X}_2=\frac{1}{n_2}\sum_{i=1}^{n_2}X_{2i}$ はそれぞれ X_{1i}，X_{2i} の標本平均である。σ_1^2，σ_2^2 が既知であれば(1)式の統計量を利用して仮説検定を行えるが，多くの場合 σ_1^2，σ_2^2 は未知であるため(1)式の実現値を計算することができないため，(1)式に基づいて仮説検定を行うこともできない。そこで(1)式の σ_1^2，σ_2^2 を不偏推定量 $s_1^2=\frac{1}{n_1-1}\cdot\sum_{i=1}^{n_1}(X_{1i}-\bar{X}_1)^2$，$s_2^2=\frac{1}{n_2-1}\sum_{i=1}^{n_2}(X_{2i}-\bar{X}_2)^2$ で置き換えた

$$T_W=\frac{\bar{X}_1-\bar{X}_2}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}} \qquad (2)$$

を用いることを考えよう。1つの正規母集団の平均に関する検定を行う場合には，未知の分散を不偏推定量で置き換えると帰無仮説のもとで t 分布にしたがうことがよく知られているが，(2)式で与えられる T_W は，帰無仮説が正しかったとしても t 分布にしたがわない。後にみるように，T_W の分布は，帰無仮説のもとでも未知パラメータである σ_1^2，σ_2^2 に依存する。このため，2つの母集団の分散が異なり，かつ未知である場合には，平均の差に関する検定を厳密な有意水準で行うことはできない[1)]。このような，2つの母集団の分散が未知で異なる可能性がある場合の，平均の差に関する検定に関わる問題はベーレンス・フィッシャー問題と呼ばれる。

上記のように，T_W の分布は t 分布ではないが，次のように考えることにより仮説検定を行うことができる。まず，$n_1\leq n_2$ であるとすれば，帰無仮説のもとで

$$P(|T_W|\geq t_{\alpha/2}(n_1+n_2-2))>\alpha,\ P(|T_W|\geq t_{\alpha/2}(n_1-1))<\alpha \qquad (3)$$

となることが知られている。ただし $t_\alpha(m)$ は自由度 m の t 分布の上側 $100\times\alpha\%$ 点である。

したがって，T_W の実現値 t_W が得られた時，$|t_W|>|t_{\alpha/2}(n_1-1)|$ ならば帰無仮説 H_0 は棄却され，$|t_W|<|t_{\alpha/2}(n_1+n_2-2)|$ ならば帰無仮説 H_0 は採択される。言い換えれば，T_m を自由度 m の t 分布にしたがう確率変数とし，T_W の実現値 t_W が得られた時の，自由度 n_1+n_2-2 および n_1-1 の t 分布に基づく p 値（p-value）をそれぞれ

$$\underline{p}=P(|T_{n_1+n_2-2}|\geq t_W),\ \overline{p}=P(|T_{n_1-1}|\geq t_W), \tag{4}$$

と定めれば，$\overline{p}\leq\alpha$ ならば帰無仮説 H_0 は棄却され，$\underline{p}\geq\alpha$ ならば帰無仮説 H_0 は採択される。

このような方法に基づくと検定結果が不確定になる場合が生じるが，Welch（1947）は，T_W の分布を自由度

$$wdf=\frac{(\sigma_1^2/n_1+\sigma_2^2/n_2)^2}{(\sigma_1^2/n_1)^2/(n_1-1)+(\sigma_2^2/n_2)^2/(n_2-1)} \tag{5}$$

の t 分布で近似することを提案した。（5）式は未知パラメータである $\sigma_1^2,\ \sigma_2^2$ を含んでいるため，実際にはこれらを不偏推定量 $s_1^2,\ s_2^2$ で置き換えた

$$\widehat{wdf}=\frac{(s_1^2/n_1+s_2^2/n_2)^2}{(s_1^2/n_1)^2/(n_1-1)+(s_2^2/n_2)^2/(n_2-1)} \tag{6}$$

である t 分布を用いて近似的な検定を行うことができる。この方法が実際の応用では最もよく用いられている方法であり，ウェルチの検定と呼ばれる。ウェルチの検定で用いられる自由度 $\widehat{wdf}$ については $n_1-1\leq\widehat{wdf}\leq n_1+n_2-2$ が成立しているため，p 値を求めると $\underline{p}$ と $\overline{p}$ の間の値が得られる。

本稿では，簡単化のために対立仮説が $H_1:\mu_1\neq\mu_2$ である場合，すなわち両側検定の場合を考える[2)]。このため，上記のウェルチの検定を行う場合には $W=T_w^2$ とすると，t 分布と F 分布の関係により，帰無仮説のもとで W が近似的に自由度 $(1,\ \widehat{wdf})$ の F 分布にしたがうことを利用して検定すれば良い。つまり，$F_\alpha(k,\ l)$ を自由度 $(k,\ l)$ の上側 $100\times\alpha\%$ 点であるとすると，W の実現値 w が得られた時，$w>F_\alpha(1,\ n_1-1)$ であれば帰無仮説は棄却され，$w<F_\alpha(1,\ n_1+n_2-2)$ であれば帰無仮説は採択されると考えることができる。p 値を用いる場合は $F(k,\ l)$ を自由度 $(k,\ l)$ の F 分布にしたがう確率変数として

$$\underline{p}=P(|F(1,\ n_1+n_2-2)|\geq t_W^2),\ \overline{p}=P(|F(1,\ n_1-1)|\geq t_W^2), \tag{7}$$

とすれば，（4）式を用いた場合と同様の方法により判断することができる。

2.2　p 値に関する考察

実際の分析においては，仮説検定を様々な有意水準で行うよりも p 値を求める方が有益な場合が多い。T_W の実現値 t_W が得られた時の p 値を，$H_0:\mu_1=\mu_2$ のもとでの $W=T_W^2$ の分布を用いて

$$p_W=P(W>t_W^2)=P(T_W^2>t_W^2) \tag{8}$$

とすると，$p_W<\alpha$ ならば帰無仮説は棄却され，$p_W\geq\alpha$ ならば帰無仮説は採択される。しかし，実際には σ_1^2 と σ_2^2 は既知でなければ p_W を計算することはできない。前にも述べたとおり，(4)式か(7)式を用いれば，$\bar{p}<\alpha$ ならば帰無仮説 H_0 は棄却され，$\underline{p}\geq\alpha$ ならば帰無仮説 H_0 は採択される。しかし，$\bar{p}$ と $\underline{p}$ の差が大きい時，検定が不確定になる可能性が高くなる。そこで，いくつかの n_1, n_2 の値に対して，$\bar{p}-\underline{p}$ の最大値を計算したものが表1である。数値の下の括弧内の値は，$\bar{p}-\underline{p}$ が最大になる t_W^2 の値である。

表1 $\bar{p}-\underline{p}$ の最大値

n_1	n_2-n_1						
	5	10	20	30	50	100	∞
5	0.03681 (2.936)	0.05144 (2.850)	0.06177 (2.790)	0.06586 (2.766)	0.06943 (2.745)	0.07230 (2.729)	0.07528 (2.712)
10	0.01046 (2.665)	0.01707 (2.633)	0.02324 (2.604)	0.02618 (2.590)	0.02903 (2.576)	0.03154 (2.564)	0.03440 (2.551)
20	0.00285 (2.539)	0.00527 (2.528)	0.00822 (2.515)	0.00994 (2.508)	0.01186 (2.500)	0.01383 (2.491)	0.01644 (2.480)
30	0.00131 (2.497)	0.00256 (2.492)	0.00428 (2.485)	0.00541 (2.480)	0.00680 (2.474)	0.00838 (2.468)	0.01079 (2.457)
50	0.00048 (2.464)	0.00100 (2.462)	0.00179 (2.459)	0.00239 (2.456)	0.00321 (2.453)	0.00430 (2.449)	0.00637 (2.440)
100	0.00012 (2.439)	0.00027 (2.439)	0.00051 (2.438)	0.00072 (2.437)	0.00106 (2.435)	0.00159 (2.433)	0.00313 (2.427)

表1から n_1 がある程度大きければ $\bar{p}$ と $\underline{p}$ に大きな差はないものの，n_1 が小さく，n_1 と n_2 の差が大きい場合には $\bar{p}$ と $\underline{p}$ の差が比較的大きくなることがわかる。$n_1=5, n_2-n_1=30$，つまり $n_2=35$ の時を例にとると，$t_W^2=2.766$ に対して $\bar{p}-\underline{p}=0.06586$ となる。この時，表には記載していないが，$\bar{p}=0.17162, \underline{p}=0.10576$ となる。したがって，$t_W^2=2.766$ であれば $\underline{p}>0.10$ であるから帰無仮説は有意水準0.10で採択されると判断できる。しかし，t_W^2 が2.766から増加すると，$\bar{p}$ と $\underline{p}$ がともに減少し，$\underline{p}<0.10<\bar{p}$ となり得ることは容易に推測できるだろう。このように，$\bar{p}$ と $\underline{p}$ に基づいて検定を行うならば，伝統的によく用いられる有意水準0.10の検定を行った場合でも，検定結果が確定しないという事態が起こり得る。

このように検定結果が確定しないという事態が生じるのは，W の分布関数が未知パラメータに依存しているため，正確な p 値が計算できないためである。次に W の分布関数を導出し，ウェルチによる近似法の特性を分析しよう。

2.3 検定統計量の分布関数

Appendixで示されているように，n_1, n_2 が奇数の時，$W=T_W^2$ の分布関数は，以下の式で

与えられる。

$$F_W(w)=\mathrm{P}(W\leqq w)$$
$$=\sum_{j=0}^{\frac{n_1-1}{2}-1}\alpha_j(2a)^{j+1}\Gamma(j+1)\sum_{i=0}^{\infty}w_i(\lambda)I_{w_a}(i+1/2,\ j+1)$$
$$+\sum_{j=0}^{\frac{n_2-1}{2}-1}\beta_j(2b)^{j+1}\Gamma(j+1)\sum_{i=0}^{\infty}w_i(\lambda)I_{w_b}(i+1/2,\ j+1)\qquad(9)$$

となる。ただし $w_a=aw/(aw+1)$, $w_b=bw/(ab+1)$,

$$a=\frac{\sigma_1^2/(n_1(n_1-1))}{\sigma_1^2/n_1+\sigma_2^2/n_2},\ b=\frac{\sigma_2^2/(n_2(n_2-1))}{\sigma_1^2/n_1+\sigma_2^2/n_2},$$

$$\alpha_j=\frac{(-1)^{(n_1-3)/2-j}\left(\frac{1}{2b}-\frac{1}{2a}\right)^{j-(n_1+n_2)/2+2}{}_{\frac{n_1+n_2}{2}-j-3}C_{\frac{n_1-3}{2}-j}}{(2a)^{(n_1-1)/2}(2b)^{(n_2-1)/2}j!},$$

$$\beta_j=\frac{(-1)^{(n_2-3)/2-j}\left(\frac{1}{2a}-\frac{1}{2b}\right)^{j-(n_1+n_2)/2+2}{}_{\frac{n_1+n_2}{2}-j-3}C_{\frac{n_2-3}{2}-j}}{(2a)^{(n_1-1)/2}(2b)^{(n_2-1)/2}j!},$$

$\lambda=(\mu_1-\mu_2)^2/\left\{\left(\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}\right)\right\}$, $w_i(\lambda)=\exp(-\lambda/2)(\lambda/2)^i/i!$ は平均 $\lambda/2$ のポアソン分布の確率関数，${}_kC_l=\frac{k!}{(k-l)!l!}$, $\Gamma(a)$ はガンマ関数

$$\Gamma(a)=\int_0^{\infty}x^{a-1}e^{-x}dx,\ (a>0),\qquad(10)$$

$I_c(a,\ b)$ は

$$I_c(a,\ b)=\frac{1}{B(a,\ b)}\int_0^c t^{a-1}(1-t)^{b-1}dt,\ (a,\ b>0)$$

$$B(a,\ b)=\int_0^c t^{a-1}(1-t)^{b-1}dt=\frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

で定義される不完全ベータ関数比である。(9)を w で微分すれば W の確率密度関数 $f_W(w)$ が得られる。

(9)式から，前述の通り，W の分布は帰無仮説 $\mathrm{H}_0:\mu_1=\mu_2$ のもとでも $\sigma_1^2,\ \sigma_2^2$ に依存することがわかる。このため，$\sigma_1^2,\ \sigma_2^2$ が未知である場合，T_W を用いて厳密な有意水準の検定を行うことはできない。

2.4　近似に関する考察

先ほど見たように，$\bar{p}$ と $\underline{p}$ を用いた検定では，伝統的によく用いられている有意水準である0.10を用いた場合でも，検定結果が不確定になる可能性がある。そこで次に Welch (1947) で提案された近似である(5)式の wdf を自由度に持つ t 分布を用いて検定を行った場合について考えよう。ただし，wdf は未知パラメータ $\sigma_1^2,\ \sigma_2^2$ を含むので，この方法は実行可能で

表2　$n_1=5,\ n_2=15$ の場合の分布関数 $F(c)$

	$(\sigma_1^2,\ \sigma_2^2)=(1,\ 5)$		$(\sigma_1^2,\ \sigma_2^2)=(5,\ 1)$				
c	True	$F(1,\ wdf)$	True	$F(1,\ wdf)$	$F(1,\ n_1-1)$	$F(1,\ n_1+n_2-2)$	$\chi^2(1)$
0.0	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
0.2	0.33932	0.33923	0.32542	0.32478	0.32213	0.33994	0.34528
0.4	0.46403	0.46390	0.44340	0.44248	0.43856	0.46496	0.47291
0.6	0.55019	0.55003	0.52425	0.52309	0.51818	0.55137	0.56142
0.8	0.61573	0.61553	0.58544	0.58406	0.57835	0.61710	0.62891
1.0	0.66790	0.66767	0.63405	0.63247	0.62610	0.66944	0.68269
1.2	0.71057	0.71032	0.67382	0.67206	0.66515	0.71224	0.72668
1.4	0.74612	0.74584	0.70704	0.70511	0.69776	0.74788	0.76328
1.6	0.77613	0.77582	0.73523	0.73313	0.72542	0.77797	0.79410
1.8	0.80172	0.80139	0.75942	0.75717	0.74918	0.80361	0.82029
2.0	0.82372	0.82337	0.78039	0.77801	0.76980	0.82564	0.84270
2.2	0.84275	0.84239	0.79873	0.79621	0.78784	0.84469	0.86199
2.4	0.85932	0.85894	0.81487	0.81223	0.80374	0.86126	0.87866
2.6	0.87381	0.87342	0.82917	0.82642	0.81784	0.87574	0.89314
2.8	0.88653	0.88613	0.84190	0.83905	0.83042	0.88845	0.90574
3.0	0.89775	0.89733	0.85330	0.85035	0.84170	0.89963	0.91674
3.2	0.90767	0.90725	0.86354	0.86051	0.85185	0.90952	0.92636
3.4	0.91647	0.91604	0.87278	0.86967	0.86103	0.91828	0.93480
3.6	0.92429	0.92386	0.88115	0.87796	0.86936	0.92606	0.94222
3.8	0.93127	0.93084	0.88875	0.88550	0.87695	0.93299	0.94875
4.0	0.93751	0.93707	0.89568	0.89237	0.88388	0.93918	0.95450
4.2	0.94310	0.94266	0.90201	0.89865	0.89023	0.94471	0.95958
4.4	0.94811	0.94767	0.90781	0.90440	0.89607	0.94968	0.96406
4.6	0.95262	0.95218	0.91313	0.90969	0.90144	0.95413	0.96803
4.8	0.95668	0.95625	0.91804	0.91456	0.90640	0.95814	0.97154
5.0	0.96035	0.95992	0.92256	0.91905	0.91099	0.96175	0.97465
5.2	0.96366	0.96324	0.92675	0.92321	0.91525	0.96501	0.97741
5.4	0.96666	0.96624	0.93062	0.92706	0.91920	0.96796	0.97986
5.6	0.96938	0.96896	0.93422	0.93063	0.92288	0.97062	0.98204
5.8	0.97184	0.97144	0.93756	0.93396	0.92631	0.97304	0.98397
6.0	0.97409	0.97368	0.94067	0.93706	0.92952	0.97523	0.98569
6.2	0.97613	0.97573	0.94357	0.93995	0.93251	0.97722	0.98722
6.4	0.97798	0.97760	0.94628	0.94265	0.93532	0.97904	0.98859
6.6	0.97968	0.97930	0.94881	0.94517	0.93796	0.98069	0.98980
6.8	0.98123	0.98086	0.95118	0.94754	0.94044	0.98219	0.99088
7.0	0.98264	0.98228	0.95340	0.94977	0.94276	0.98356	0.99185
7.2	0.98394	0.98359	0.95549	0.95185	0.94496	0.98482	0.99271
7.4	0.98513	0.98478	0.95745	0.95382	0.94703	0.98597	0.99348
7.6	0.98621	0.98588	0.95930	0.95567	0.94898	0.98702	0.99416
7.8	0.98721	0.98689	0.96103	0.95741	0.95083	0.98798	0.99478
8.0	0.98813	0.98781	0.96267	0.95906	0.95258	0.98886	0.99532
8.2	0.98897	0.98866	0.96422	0.96062	0.95424	0.98967	0.99581
8.4	0.98975	0.98945	0.96568	0.96209	0.95581	0.99042	0.99625
8.6	0.99047	0.99017	0.96706	0.96348	0.95730	0.99110	0.99664
8.8	0.99113	0.99084	0.96837	0.96480	0.95871	0.99173	0.99699
9.0	0.99173	0.99146	0.96961	0.96605	0.96006	0.99231	0.99730
9.2	0.99230	0.99202	0.97078	0.96724	0.96134	0.99285	0.99758
9.4	0.99281	0.99255	0.97189	0.96837	0.96256	0.99334	0.99783
9.6	0.99329	0.99304	0.97295	0.96945	0.96372	0.99380	0.99805
9.8	0.99374	0.99349	0.97395	0.97047	0.96483	0.99422	0.99825
10.0	0.99415	0.99391	0.97491	0.97144	0.96589	0.99461	0.99843

表 3　$n_1=5,\ n_2=15$ の場合の棄却率

	α										
$(\sigma_1^2,\ \sigma_2^2)$	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	0.95	0.99
(1, 5)	0.0970	0.1958	0.2960	0.3964	0.4970	0.5966	0.6969	0.7987	0.8985	0.9495	0.9901
(5, 1)	0.1037	0.1996	0.2971	0.3973	0.4979	0.5985	0.7003	0.7992	0.9003	0.9510	0.9900

はないことに注意が必要である。検定の p 値を求めるためには帰無仮説のもとでの分布関数を用いるので，ここでは帰無仮説のもとでの $W=T_W^2$ の分布関数，つまり(9)式において $\lambda=0$ としたものを様々な w の値に対して計算し，自由度 wdf の t 分布にしたがう確率変数を 2 乗したものの分布関数，つまり自由度 $(1,\ wdf)$ の F 分布の分布関数と比較する。

表 2 で計算されているのが任意の分布関数を $F(\cdot)$ とした時の $F(c)$ の値である。表中で「True」で表されているのが真の分布関数，つまり(9)式において $\lambda=0$ としたものの分布関数であり，$F(1,\ wdf)$ は Welch (1947) で提案された近似的な分布関数，つまり自由度 $(1,\ wdf)$ の F 分布の分布関数である。また，比較のために $\bar{p}$ と $\underline{p}$ を与える分布関数である自由度 $(1,\ n_1-1)$ と自由度 $(1,\ n_1+n_2-2)$ の F 分布の分布関数，正規分布による近似を行なった場合の分布となる自由度 1 のカイ 2 乗分布の分布関数も掲載している。

表 2 から，実際の W の分布は $(\sigma_1^2,\ \sigma_2^2)=(1,\ 5)$ の時には自由度 $(1,\ n_1+n_2-2)$ の F 分布に近く，$(\sigma_1^2,\ \sigma_2^2)=(5,\ 1)$ の時には自由度 $(1,\ n_1-1)$ の F 分布に近いことがわかる。このことから，W の分布は自由度 $(1,\ n_1-1)$ と $(1,\ n_1+n_2-2)$ の F 分布のどちらにも近づき得るので，どちらかの分布をより良い近似ということはできない。これに対し，自由度 $(1,\ wdf)$ の F 分布は，W の真の分布の非常に良い近似となっていることがわかる。また，正規近似を用いた場合，つまり自由度 1 のカイ 2 乗分布を用いた場合の近似はあまり良くない。したがって，自由度 $(1,\ wdf)$ の F 分布を用いるウェルチの検定は，σ_1^2 と σ_2^2 の値が既知であれば非常に有益な方法であると考えられる。しかしながら，通常は σ_1^2 と σ_2^2 は未知であるため，実際の応用では σ_1^2 と σ_2^2 の推定量である s_1^2 と s_2^2 の推定精度が重要となる。そこで wdf の σ_1^2 と σ_2^2 を推定量 $s_1^2,\ s_2^2$ で置き換えた $\widehat{wdf}$ を用いた場合にどのような影響があるのかを調べよう。実際に $\widehat{wdf}$ を用いた場合の分析を解析的に行うのは困難であるため，コンピューターを用いたシミュレーションを行うことにより分析する。

表 3 は，ウェルチの検定の棄却率，つまり帰無仮説のもとで発生させた標本から計算した W の実現値が，自由度 $(1,\ \widehat{wdf})$ の F 分布の上側確率が α となる点より大きくなる比率を，100000回の繰り返し実験（シミュレーション）に基づいて計算したものである。棄却率が有意水準である上側確率 α に近ければ望ましい検定が行えることを意味している。表 3 は表 2 と同様の設定で実験を行った結果である。この結果から，どちらの場合もかなり正確な検定を行える，つまり wdf を $\widehat{wdf}$ で置き換えたことによる影響は非常に小さいと考える

ことができる。

3　他のいくつかの方法との比較

本節では，ウェルチの検定とは異なる平均の差の検定をいくつか紹介し，その特性を数値計算を用いてウェルチの検定と比較する。

3.1　シェッフェの方法

前節において，一般的によく用いられているウェルチの検定が近似的な方法であることを説明したが，正規母集団において分散が未知で異なる場合でも，得られた標本を全て用いることを諦めれば[3]，近似的ではなく厳密な有意水準で平均の差の検定を行う方法がある。以下はScheffé（1970）により提案された方法でありシェッフェの方法と呼ばれる[4]。

これまでと同様 $n_1 \leq n_2$ の場合を考え，X_{2i} からランダムに n_1 個選んだものを X_{2i}^*（$i=1, 2, \ldots, n_1$）とする。この時

$$U_i = X_i - \sqrt{\frac{n_1}{n_2}} X_{2i}^* \quad i=1, 2, \ldots, n_1 \tag{11}$$

とすれば

$$U_i \sim \mathrm{N}\left(\mu_1 - \sqrt{\frac{n_1}{n_2}}\mu_2,\ \sigma_1^2 + \frac{n_1}{n_2}\sigma_2^2\right) = \mathrm{N}\left(\mu_1 - \sqrt{\frac{n_1}{n_2}}\mu_2,\ \tilde{\sigma}^2\right) \tag{12}$$

となる。ただし $\tilde{\sigma}^2 = \sigma_1^2 + n_1\sigma_2^2/n_2$ である。また，$\bar{U} = \frac{1}{n_1}\sum_{i=1}^{n_1} U_i$，$s_u^2 = \sum_{i=1}^{n_1}(U_i - \bar{U})^2/(n_1-1)$ とすれば

$$(n_1-1)s_u^2/\tilde{\sigma}^2 \sim \chi^2(n_1-1) \tag{13}$$

となる。$\bar{X}_1 - \bar{X}_2 \sim \mathrm{N}(\mu_1 - \mu_2,\ \tilde{\sigma}^2/n_1)$ であることを用いれば

$$T_S = \frac{(\bar{X}_1 - \bar{X}_2)/\sqrt{\tilde{\sigma}^2/n_1}}{\sqrt{\dfrac{(n_1-1)s_u^2}{\tilde{\sigma}^2}/(n_1-1)}} = \frac{\sqrt{n_1}(\bar{X}_1 - \bar{X}_2)}{\sqrt{s_u^2}} \tag{14}$$

は帰無仮説 $\mathrm{H}_0 : \mu_1 = \mu_2$ のもとで自由度 $n_1 - 1$ の t 分布にしたがう。これは近似的な分布ではないので，厳密な有意水準の検定を行うことができる。

この検定の問題は，X_{2i} から n_1 個の X_{2i}^* を選ぶため，分母の計算において全てのサンプルを用いていない点と，n_1 個の X_{2i}^* はランダムに選ばれるため，どれが選ばれたかによって検定結果が異なることである。後者については，$n_1 = n_2$ の場合でも

$$\begin{aligned}(n_1-1)s_u^2 &= \sum_{i=1}^{n_1}\{(X_{1i} - \bar{X}_1) - (X_{2i}^* - \bar{X}_2)\}^2 \\ &= \sum_{i=1}^{n_1}(X_{1i} - \bar{X}_1)^2 + \sum_{i=1}^{n_1}(X_{2i}^* - \bar{X}_2)^2 + 2\sum_{i=1}^{n_1}(X_{1i} - \bar{X}_1)(X_{2i}^* - \bar{X}_2)\end{aligned} \tag{15}$$

の第3項が X_{2i}^{*} がどのように並べられたか，言い換えれば，どの X_{1i} と X_{2i} の組み合わせで U_i が作成されたかに依存するため，問題が解消されることはない。しかしながら，この項を n で割れば標本相関が得られるので，X_{1i} と X_{2i} が独立であるから大きな影響は無いと考えることもできる。

これまでと同様に対立仮説が $H_1:\mu_1\neq\mu_2$ である場合，すなわち両側検定の場合には

$$W_S=T_S^2 \tag{16}$$

が帰無仮説のもとで自由度 $(1,\ n_1-1)$ の F 分布にしたがうことを用いて厳密な有意水準の検定を行える。また，対立仮説の下では W_S は自由度 $(1,\ n_1-1)$，非心度 $\lambda=(\mu_1-\mu_2)^2/\left\{\left(\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}\right)\right\}$ の非心 F 分布にしたがう。

3.2　分散が等しい場合の検定

分散の値は未知ではあるが，2つの母集団の分散が等しいことはわかっている場合，つまり $\sigma_1^2=\sigma_2^2=\sigma^2$ で σ^2 が未知の場合には，次のように考えることで厳密な有意水準で検定が行える。

まず

$$s^2=\frac{\sum_{i=1}^{n_1}(X_{1i}-\bar{X}_1)^2+\sum_{i=1}^{n_2}(X_{2i}-\bar{X}_2)^2}{n_1+n_2-2} \tag{17}$$

とすれば

$$\frac{n_1+n_2-2}{\sigma^2}s^2\sim\chi^2(n_1+n_2-2) \tag{18}$$

が成り立つ。

$\bar{X}_1-\bar{X}_2\sim N\left(\mu_1-\mu_2,\ \left(\frac{1}{n_1}+\frac{1}{n_2}\right)\sigma^2\right)$ であるから

$$T_E=\frac{(\bar{X}_1-\bar{X}_2)/\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\sigma^2}}{\sqrt{\frac{n_1+n_2-2}{\sigma^2}s^2/(n_1+n_2-2)}}=\frac{\bar{X}_1-\bar{X}_2}{\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)s^2}} \tag{19}$$

は帰無仮説 $H_0:\mu_1=\mu_2$ のもとで自由度 n_1+n_2-2 の t 分布にしたがう。このことを用いて厳密な有意水準での検定を行うことができる。$\sigma_1^2=\sigma_2^2=\sigma^2$ が真であれば，この情報を用いているため，この検定はウェルチの検定よりも高い検定力を持つことが期待できる。

対立仮説が $H_1:\mu_1\neq\mu_2$ である場合には

$$W_E=T_E^2 \tag{20}$$

が帰無仮説のもとで自由度 $(1,\ n_1+n_2-2)$ の F 分布にしたがうことを用いて厳密な有意水準

の検定を行える。また，対立仮説の下では W_E は自由度 $(1, n_1+n_2-2)$，非心度 $\lambda'=(\mu_1-\mu_2)^2/\left\{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\sigma^2\right\}$ の非心 F 分布にしたがう。

3.3　数値計算による分析

ここではウェルチの検定とシェッフェの方法，分散が等しい場合の検定である W_E を用いた検定の検定力を数値計算により比較する。数値計算においては，$\sigma_1^2=\sigma_2^2=\sigma^2$ である場合を考える。これは，本当は2つの母集団の分散は等しいが，分析者がそのことを知らないためウェルチの検定を行なった場合，どの程度検定結果に影響があるかを調べるためである。検定力の計算においては，ウェルチの検定とシェッフェの方法では λ，分散が等しい場合の W_E を用いた検定では λ' の値が必要になるが，$\sigma_1^2=\sigma_2^2=\sigma^2$ の時には $\lambda=\lambda'$ となるので $\mu_1-\mu_2$ に値を与えることで様々な λ に対する検定力を計算する。また，検定力を比較する際には，有意水準が正確になるような棄却域を用いて比較を行う必要があるため，ウェルチの検定については帰無仮説が正しい場合，つまり $\lambda=0$ の場合の(9)式を用いて棄却点を求めて比較を行う。

数値計算で求めた検定力が表3である。表3から n_1 が固定された場合，n_2 が大きい方がどの方法でも検定力が高くなるのがわかる。また，2つの母集団の分散が等しい場合，W_E に基づく検定力が確かに一番大きい。しかしながら，$n_2=15$ の場合を見ると，ウェルチの検定も W_E を用いた検定にかなり近い検定力を持っている。このことは，本当は2つの母集団の分散が等しかったとしても，2つの母集団の分散が等しいか分析者がわからない場合には，ウェルチの検定を行なっても大きな弊害が無いことを意味している。また，シェッフェの方法については，3つの方法の中で最も検定力が小さい。$n_2=55$ になると，3つの方法の検定力の差は $n_2=15$ の場合に比べて広がる傾向があることがわかる。しかしながら，シェッフェの方法の検定力は3つの方法の中で最も小さく，現実的な問題では2つの母集団の分散が等しいと仮定できるケースはあまり多く無いことと，ウェルチの検定の検定力が，W_E を用いた検定と比較して極端に低いわけでは無いことを考えれば，ウェルチの検定を用いるのは妥当であるといえるだろう。

4　結　　論

本稿では，2つの正規母集団の平均の差を検定する方法としてウェルチの検定と，関連するいくつかの方法について分析を行った。まず，ウェルチの検定は，近似的な方法ではあるものの，非常に高い近似精度を持つ事が数値計算とシミュレーション結果により確認された。また，分散が等しいことを必要としないウェルチの検定を行った場合，仮に2つの母集団の

表 4　ウェルチの検定，シェッフェの方法，等分散の場合の検定力

	$n_1=5,\ n_2=15$			$n_1=5,\ n_2=55$		
$\mu_1-\mu_2$	Welch	Scheffé	W_E	Welch	Scheffé	W_E
0.0	0.1000	0.1000	0.1000	0.1000	0.1000	0.1000
0.1	0.1054	0.1045	0.1059	0.1060	0.1055	0.1076
0.2	0.1215	0.1179	0.1235	0.1240	0.1219	0.1302
0.3	0.1481	0.1400	0.1526	0.1535	0.1488	0.1676
0.4	0.1847	0.1706	0.1927	0.1941	0.1859	0.2187
0.5	0.2306	0.2090	0.2429	0.2446	0.2322	0.2821
0.6	0.2848	0.2545	0.3021	0.3037	0.2865	0.3555
0.7	0.3457	0.3061	0.3684	0.3695	0.3474	0.4359
0.8	0.4116	0.3626	0.4396	0.4397	0.4130	0.5195
0.9	0.4803	0.4225	0.5132	0.5119	0.4812	0.6025
1.0	0.5497	0.4843	0.5866	0.5836	0.5498	0.6813
1.1	0.6175	0.5464	0.6571	0.6524	0.6168	0.7526
1.2	0.6818	0.6072	0.7225	0.7162	0.6801	0.8144
1.3	0.7409	0.6653	0.7812	0.7736	0.7384	0.8656
1.4	0.7936	0.7194	0.8320	0.8235	0.7905	0.9061
1.5	0.8392	0.7688	0.8745	0.8657	0.8357	0.9368
1.6	0.8776	0.8127	0.9089	0.9003	0.8739	0.9591
1.6	0.8776	0.8127	0.9089	0.9003	0.8739	0.9591
1.8	0.9339	0.8834	0.9560	0.9489	0.9304	0.9847
1.9	0.9531	0.9104	0.9707	0.9648	0.9500	0.9912
2.0	0.9675	0.9324	0.9811	0.9763	0.9649	0.9951
2.1	0.9780	0.9500	0.9882	0.9845	0.9759	0.9974
2.2	0.9855	0.9636	0.9928	0.9901	0.9838	0.9987
2.3	0.9907	0.9740	0.9958	0.9938	0.9894	0.9994
2.4	0.9941	0.9818	0.9976	0.9962	0.9932	0.9997
2.5	0.9964	0.9874	0.9987	0.9978	0.9957	0.9999
2.6	0.9978	0.9915	0.9993	0.9987	0.9974	0.9999
2.6	0.9978	0.9915	0.9993	0.9987	0.9974	0.9999
2.8	0.9993	0.9963	0.9998	0.9996	0.9991	1.0000
2.9	0.9996	0.9977	0.9999	0.9998	0.9995	1.0000

分散が等しかったとしても検定力はあまり低下しないことが示された。さらに，母集団の分散が異なった場合でも厳密な検定を行えるシェッフェの方法と比較すると，ウェルチの検定の方が検定力が高い事がわかった。上記の結果から，母集団の分散が等しいかわからない現実の状況では，ウェルチの検定を行うのが妥当であると考える事ができる。

本稿の分析で考えている母集団は正規母集団に限定されている。実際の分析では母集団が正規母集団かどうかさえ未知である場合が多い。このような場合でも，標本がある程度大きければ本稿の分析は有効であるとは考えられるが，標本が大きい場合にはそもそも漸近分布である正規分布を用いた近似で十分かもしれない。これに対し，母集団が正規母集団ではな

く標本が小さい場合には，平均の差を検定するウェルチの検定よりもマン・ホイットニーの U 検定［Mann and Whitney（1947）］や Brunner-Munzel 検定［Brunner and Munzel（2000）］を用いて分布の差を検定する方が理想的かもしれない。しかし，Rasch et al.（2011）のように，分布が未知の場合でもマン・ホイットニーの U 検定を行うよりウェルチの検定を行った方が良いというシミュレーション結果もある。また，本稿においては1次元の確率変数の平均の差について分析を行ったが，Wang and Xu（2022），Zhang et al.（2021）のように，高次元確率変数におけるベーレンス・フィッシャー問題に関する分析も行われている。このように母集団の分布に依存しないノンパラメトリックな場合や高次元の場合への拡張など，ベーレンス・フィッシャー問題は現在でも重要な研究トピックといえるだろう。

Appendix

ここでは（9）で与えられる W の分布関数 $F_W(w)$ を導出する。まず，W は

$$W=\frac{X^2}{aY+bZ}, \tag{21}$$

$$X=\frac{(\bar{X}_1-\bar{X}_2)^2}{\sigma_1^2/n_1+\sigma_2^2/n_2}, \tag{22}$$

と書くことができる。ただし $Y=(n_1-1)s_1^2/\sigma_1^2\sim\chi^2(n_1-1)$, $Z=(n_2-1)s_2^2/\sigma_2^2\sim\chi^2(n_2-1)$ であり，X は非心度 $\lambda=(\mu_1-\mu_2)^2/\left\{\left(\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}\right)\right\}$ の非心カイ2乗分布にしたがい，X, Y, Z は互いに独立である。ここで，Ray and Pitman（1961）で導出されているように，W の分母 $V=aY+bZ$ の確率密度関数は

$$f(v)=e^{-\frac{v}{2a}}\sum_{j=1}^{\frac{n_1-1}{2}-1}\alpha_j v^j+e^{-\frac{v}{2b}}\sum_{j=1}^{\frac{n_2-1}{2}-1}\beta_j v^j \tag{23}$$

で与えられる。X, Y, Z が独立であるから V と X も独立であるため，Y と X の同時確率密度関数は $f(v)f(x)$ で与えられ，$F(w)$ は

$$\begin{aligned}F(w)&=\iint_{x/v<w}f(v)f(x)dx\,dv\\&=\sum_{j=1}^{\frac{n_1-1}{2}-1}\alpha_j\sum_{i=1}^{\infty}w_i(\lambda)\frac{1}{2^{i+1/2}\Gamma(i+1/2)}\iint_{x/v<w}x^{1/2+i-1}e^{-\frac{x}{2}-\frac{v}{2a}}v^j dx\,dv\\&\quad+\sum_{j=1}^{\frac{n_2-1}{2}-1}\beta_j\sum_{i=1}^{\infty}w_i(\lambda)\frac{1}{2^{i+1/2}\Gamma(i+1/2)}\iint_{x/v<w}x^{1/2+i-1}e^{-\frac{x}{2}-\frac{v}{2b}}v^j dx\,dv\end{aligned} \tag{24}$$

となる。ここで $u=x/v$ と変数変換を行えば，(24)式の第1項の積分は

$$\int_0^{\infty}\int_0^{w}v^{1/2+i+j}u^{1/2+i-1}e^{-\frac{au+1}{2a}v}du\,dv \tag{25}$$

となる。次に $t=\frac{ua+1}{2a}v$ と変数変換を行うと，この積分は

$$\Gamma(3/2+i+j)(2a)^{3/2+i+j}\int_0^{w}\left(\frac{1}{au+1}\right)^{3/2+i_j}u^{1/2+j-1}du \tag{26}$$

となる。最後に $r=\frac{au}{au+1}$ と変数変換を行えば，(26)式の積分は

$$a^{-\frac{1}{2}-i}B\left(\frac{1}{2}+i,\ j+1\right)I_{w_a}(1/2+i,\ j+1) \tag{27}$$

となる。(24)式の第2項についても同様の計算を行い，若干の計算を行うことで(9)式が得られる。

注

1）十分統計量を用いて検定統計量を作成することを考えた時，帰無仮説のもとでの検定統計量の分布が $\sigma_1^2,\ \sigma_2^2$ に依存しないようにできないことが知られている。竹内（1963）も参照。

2）片側検定の場合への拡張は容易であるが，数値計算の結果を確認しやすくするためにも両側検定を用いる。

3）厳密には，シェッフェの方法は十分統計量である $\bar{X},\ \bar{Y},\ s_{12},\ s_{22}$ を全て用いた検定ではない。

4）竹内(1963)も参照。

参 考 文 献

Brunner, E. and U. Munzel (2000). "The Nonparametric Behrens-Fisher Problem: Asymptotic Theory and a Small-Sample Approximation". *Biometrical Journal: Journal of Mathematical Methods in Biosciences* 42(1), pp. 17–25.

Mann, H. B. and D. R. Whitney (1947). "On a Test of Whether One of Two Random Variables is Stochastically Larger than the Other". *The Annals of Mathematical Statistics*, pp. 50–60.

Rasch, D., K. D. Kubinger, K. Moder (2011). "The Two-Sample T Test: Pre-Testing its Assumptions Does Not Pay Off". *Statistical papers* 52(1), p. 219.

Ray, W. and A. Pitman (1961). "An Exact Distribution of the Fisher–Behrens–Welch Statistic for Testing the Difference between the Means of Two Normal Populations with Unknown Variances". *Journal of the Royal Statistical Society: Series B* (*Methodological*) 23(2), pp. 377–384.

Scheffé, H. (1970). "Practical Solutions of the Behrens-Fisher Problem". *Journal of the American Statistical Association* 65(332), pp. 1501–1508.

Wang, R. and W. Xu (2022). "An Approximate Randomization Test for the High-Dimensional Two-Sample Behrens–Fisher Problem under Arbitrary Covariances". *Biometrika* 109(4), pp. 1117–1132.

Welch, B. L. (1947). "The Generalization of 'Student's' Problem when Several Different Population Variances are Involved". *Biometrika* 34(1–2), pp. 28–35.

Zhang, J-Ting, Zhou, B., Guo, J. and Zhu, T. (2021). "Two-Sample Behrens–Fisher Problems for High-Dimensional Data: A Normal Reference Approach". *Journal of Statistical Planning and Inference* 213, pp. 142–161.

竹内啓（1963）．数理統計学：データ解析の方法．東洋経済新報社.