

t 検定における標本サイズ決定に関する一考察

A Study on Determining Sample Size for a t-test.

2026 年 3 月 30 日

富沢日出夫	Hideo Tomizawa
福岡 宣行	Nobuyuki Fukuoka

t 検定における標本サイズ決定に関する一考察

A Study on Determining Sample Size for a t-test.

富沢日出夫¹

Hideo Tomizawa

福岡 宣行²

Nobuyuki Fukuoka

Abstract

We examined the sample size required for a t-test in relation to the central limit theorem and the shape of the population distribution. In addition, we performed simulations from the perspective of effect size. Results indicated that the rejection rate is highly dependent on the shape of the population distribution. Determining the exact number of samples required was difficult, but we estimated that a sample size of approximately 20 to 30 samples is necessary. If a test must be performed with a small sample, we believe that a nonparametric test should be used.

1. はじめに

独立した2標本の平均値の差を検定する際、広く用いられる統計的仮説検定はt検定であろう。手軽に利用できる表計算ソフトが普及し、その機能の一つとしてt検定が標準装備されていることや、近年のDXブームも相まって、必ずしもデータ解析の専門家ではない実務家でも統計的仮説検定への関心が高まっている。以前に比べてt検定が様々な場面で活用される時代になったともいえよう。このようなときに実務家を悩ませるのが標本サイズである。筆者等も永く実務家に対する統計解析の指導を行ってきたが、検定を行う際に必要な標本サイズに関しては幾度となく質問が寄せられた経験がある。正確な結果を出力できる標本サイズは多くの実務家にとっての関心事である。

しかし、一般に普及している解説書、専門書によると標本サイズに関しては「十分な大きさ」と記されているのがほとんどであり、具体的な目安を示している例は少ない。また、具体的な目安が示されていたとしても、その根拠が明示している例は見当たらない。例えば松原ら（1991）は30程度が目安であると解説しているが、その根拠までは触れていない。そこ

2026年3月30日 受理

¹ 産業能率大学経営学部 ² 学校法人産業能率大学経営管理研究所

で、本論では t 検定の出力を正確に得ることができる標本サイズを決めるための基本的な理論と考え方の整理を行い、実務において t 検定を実行する際の必要最小限の標本サイズの目安を提案したい。

2. 論点の整理

そもそも t 検定は正規分布を基盤とした Z 検定における小標本問題に対応するために W. S. Gosset (Student, 1908) が開発した検定方法である。

小標本問題とは標本サイズが小さいと Z 検定において α エラーを過小評価するものである。 α とは一般的には有意水準または危険率とも呼ばれるが、平均値に差が無いにも関わらず無仮説を棄却してしまう確率のことであり、通常 0.05 とされる。例えば、標本サイズ 10 程度の小標本で Z 検定を行うと、平均値に差が無い標本であったとしても、棄却される確率が 0.05 以上となってしまうことがある。これが小標本問題と呼ばれている現象である。

そもそも、サイズ n の標本が母平均 μ 、母標準偏差 σ からサンプリングされたものか否かを Z 検定により確かめる検定統計量は、標本平均を $\bar{x}$ とすると図-2 左側のように計算できる。しかし、分析を行う上で、母標準偏差は不明ことが多い。そのため、 σ ではなく標本標準偏差 s で代用されることも多い。ところが、この標本標準偏差は不偏量ではなく (平松ら, 2003)、小標本の場合は過小評価されやすくなる。その結果、検定統計量である Z 値が過大になり、 α エラーが過小評価される (実際の棄却率が 0.05 を上回る) のである。

標本標準偏差が負の偏りを持つことを示すために、統計解析ツール R の乱数生成関数 (`rnorm (n=3, mean=0, sd=1)`) を用いて発生させた乱数の標準偏差を計算することを 10000 回繰り返し、得られた標準偏差のヒストグラムを図-1 に示した。得られた標準偏差のピーク

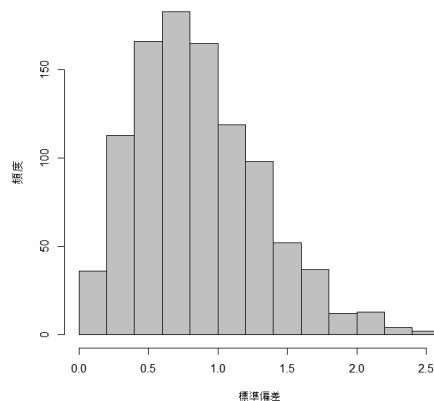


図-1 シミュレーションで得られた標準偏差の分布
(計算に用いたスクリプトは参考-1 に示す)

は設定した標準偏差 1 よりも左側に偏っている。すなわち負の偏りがあることが確認できる。結果として小標本の場合は、Z 検定は検定統計量が偏ってしまう可能性が高いことがわかる。

この問題に対応するために、W.S.Gosset は、自由度のみに依存する t 分布を開発した。標本サイズに依存しないのであれば、検定統計量である t 値は小標本でも頑健であることが期待できる。この t 値は、正規分布における標本平均の標準化における σ を標本標準偏差で代用（student 化）したものである（Student, 1908）。そのため、t 検定の対象は正規母集団を前提としており、非正規母集団の場合には検定結果が不安定になる可能性も考えられる。

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \quad \longrightarrow \quad t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

$\bar{x}$: 標本平均	$\bar{x}$: 標本平均
μ : 母平均	μ : 母平均
σ : 母標準偏差	s : 標本標準偏差
n : 標本サイズ	n : 標本サイズ

図－２ 正規分布の student 化による t 分布

しかし、独立 2 群の平均値の差の検定であれば、検定統計量の算出に用いるデータは平均値であり、中心極限定理に従って概ね正規分布に従うことが期待される。中心極限定理とは、平均をとる変数の数が多くなるにしたがって、その平均値（または総和）は正規分布に従うというものである（Walker, 1969）。したがって、t 検定において、標本の正規性に関してはそれほど神経質にならなくても良いとする意見もある（中原, 2022）。しかし、どの程度の標本サイズで正規分布となるのか、分布の形状を問わず正規性を示す標本サイズは一定なのか、具体的な検証を行った例は見当たらない。いずれにしても、t 検定は正規分布を仮定したパラメトリック検定であり、標本の正規性が担保されなければ正しい検定結果を出力することはできないだろう。

そこで、本研究では意味のある t 検定の出力を得られる標本サイズを考えていきたい。具体的には、標本平均が中心極限定理に従い正規性を示すのであれば、それはどの程度のサイズが必要なのか、そして、それを分布の形と関連させながら整理していく。

また、統計的検定は効果量、検定力、標本サイズ、有意水準の 4 要因で精度が決まってくるとされている（水本ら, 2010）。一般的には有意水準は 0.05、検定力は 0.8 が標準とされている。本稿ではその条件の下で、効果量と標本サイズの関係も検討したい。そして、t 検定を実行する上での標本サイズの捉え方について総合的に考察していきたい。

3. t 検定のモンテカルロシミュレーション

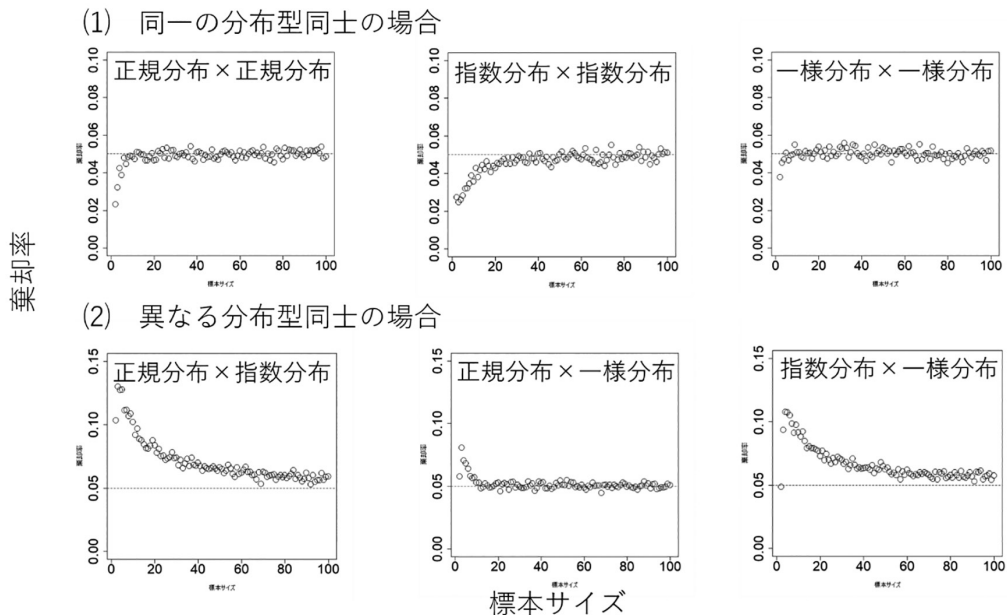
まずは、母集団分布の形状と標本サイズがt検定の頑健性にどのような影響を与えるか調べてみたい。ここでは既知の確率分布から乱数を発生させるモンテカルロシミュレーションにより検討を加えた。乱数を発生させる分布は正規分布、指数分布、一様分布の3種類とした。それぞれ期待値が等しくなるように引数を設定し（正規分布：mean=10、指数分布：rate=1/10、一様分布：min=5 max=15）、2～100の範囲で標本数を変化させ、Welch 検定を行い、p 値を求めた。この計算を標本数毎に10000回繰り返し、 $p < 0.05$ となる割合を求め棄却率とした。そもそも、期待値が等しい標本であるため、棄却率は0.05となるはずである。このシミュレーションでは、標本サイズと母集団の形状の組み合わせがWelch 検定の頑健性にどのような影響を与えるかを確認するものである。比較する2群の分布の組み合わせは①正規分布同士②指数分布同士③一様分布同士④正規分布と指数分布⑤正規分布と一様分布⑥指数分布と一様分布の6パターンでシミュレーションを行った。計算に用いたスクリプトは参考2のとおりである（ここではrateパラメーターが等しい指数分布からのシミュレーションのみの場合を例として挙げる）。

標本サイズ毎の棄却率を図－3に示す。同一分布型からの標本同士の場合（図－3の(1)）は、総じて小標本では棄却率が0.05を下回っていた。具体的には、正規分布×正規分布、一様分布×一様分布の結果は標本サイズ10前後までは0.05未満であったが、10を超えると概ね0.05前後で安定していた。指数分布×指数分布の場合は標本サイズ20前後までは0.05を下回っていた。つまり、同じ分布型から得られた標本の場合、標本サイズが小さいと帰無仮説を棄却し難い傾向であることが解った。これは、標本サイズが小さくなると標準誤差が大きくなり、t統計量が過小となりやすく、結果として帰無仮説を棄却し難い状態となっていると考えられる。つまり、標本サイズが小さいことによる検定力の低下が生じているのだろう。これらのことから、同じ分布型であれば、安定的な有意水準を維持するためには、少なくとも20程度の標本サイズが必要であることを示唆していると考えられる。このことは、標本サイズ20未満の場合は、対立仮説が正しいにも関わらず、帰無仮説が正しいとしてしまう第2種の過誤を生じさせてしまう可能性があることも示している。

一方、異なる分布型からの標本の場合（図－3の(2)）は、総じて小標本では棄却率が0.05を上回っていた。正規分布×指数分布、指数分布×一様分布の場合は、両者とも標本サイズ20前後あたりまでは急激に棄却率が低下するが、その後は鈍減であり、標本サイズ100でも0.05に達しなかった。正規分布×一様分布の場合は概ね標本サイズ20前後まで急減するが、その後は漸減し、標本サイズ60程度で0.05付近に安定した。同一分布型と異なり小標本で棄却率が大きくなっているのは興味深い。小標本なので、t統計量を算出する際の標準誤差は同一型分布同様に大きくなると考える。ここから、棄却率が大きくなっているということは

単純に平均値の差が広がっているのではないかと推測できる。異なる分布型からサンプリングした標本であり、期待値が等しい分布であったとしても、小標本の場合は同一分布に比べ標本平均のバラつきが大きくなり、平均値の差が大きくなったものと推測できる。これらのことから、母集団の分布型が異なる場合は、帰無仮説が正しいにもかかわらず対立仮説を正しいと判定する第1種の過誤が発生しやすくなることを示唆している。

このシミュレーションの結果は、t検定は母集団分布の形状と標本サイズにp値が大きく影響を受けることを示している。これは、中心極限定理による正規性の発現の仕方が母集団の形状により異なることが原因かもしれない。以降でこの点を確認してみたい。



図ー 3 標本サイズを変化させたときの非正規母集団からの2標本の棄却率の変化

4. 標本サイズと中心極限定理

中心極限定理を考えた場合であっても、十分な大きさの標本の平均は正規分布を示すとの説明は多いが（例えば、Bishop, 2012、松原ら, 1991、西内, 2014）、正規性を示すには具体的にどの位の標本サイズが最低限必要なのか言及している例は見当たらない。また、分布の形状の影響について言及している例も見当たらない。ここでは、非正規分布である母集団の形状を取り上げ、標本平均がどの程度のサイズから正規分布に近似するのか確認したい。

4.1 正規性の確認

中心極限定理とは、平均 μ 、分散 σ^2 の母集団から標本サイズ n のサンプリングを行い、都度標本平均を求めると、 $n \rightarrow \infty$ とすると、その標本平均は平均 μ 、分散 σ^2/n の正規分布に近似するというものである。

正規性を確かめる方法はいくつかある。代表的なものとしては、Kolmogorov-Smirnov 検定、Shapiro-Wilk 検定といった検定による検証、そして Q-Q プロットを描き視覚的に確認する方法、そして頻度分布を描くことで視覚的に確認する方法などが知られている。本節では、非正規分布として指数分布と一様分布を取り上げ、以下でシミュレーションした。

4.2 正規性の検定～ Shapiro-Wilk 検定

Shapiro-Wilk 検定は、アメリカの統計学者 Samuel Sanford Shapiro とカナダの統計学者 Martin Bradbury Wilk によって開発された手法であり、統計解析ツール R にも標準で関数が実装されている極めて一般的な検定方法である。この検定では、与えられたデータが正規分布からの標本であるとの帰無仮説を検証するものであり、検定統計量は以下で求められる。

$$W = \frac{(\sum_{i=1}^n a_i x_{(i)})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

x_i : i 番目の順序統計量。標本の中で i 番目に小さな値。

$\bar{x}$: 標本平均

a_i : 以下で与えられる

$$(a_1, a_2, \dots, a_n) = \frac{m^T V^{-1}}{(m^T V^{-1} V^{-1} m)^{\frac{1}{2}}}$$

$m = (m_1, m_2, \dots, m_n)$

m_i : 標準正規分布からサンプリングされた標本の順序統計量の期待値

Shapiro-Wilk 検定の基本的な考え方は Q-Q プロットと類似している。すなわち、得られたデータの順序統計量と、正規分布から得られる順序統計量の期待値に差が大きくなると検定統計量の W の値が大きくなる。正規分布から得られた標本であるとする帰無仮説の下で検定統計量 W が得られる p 値を算出し、0.05 未満の場合は正規分布から得られた標本であることを棄却する。

今回は、指数分布 (rate=1/10)、一様分布 (min=5, max=15) からそれぞれの標本サイズで乱数を発生させ平均値を取るという試行を 100 回繰り返すことで、それぞれの標本サイズにおける標本平均を 100 個得た。それぞれの標本サイズ毎に統計解析ツール R に実装されてい

る Shapiro-Wilk 検定を実行する関数 (shapiro.test) を用いて検定をおこなった。それぞれの標本サイズにおいて棄却率 ($p\text{-value} < 0.05$) となった割合を図-4 及び図-5 に示す。すなわち、この値が大きければ正規性が棄却される割合が大きいことを示している。また、計算に用いた R スクリプトは参考3 に示す (ここでは指数分布についてのみ示す)。

指数乱数の場合、概ね棄却率は 0.1 前後であったが、一様乱数は 0.01 前後であった。指数分布からの標本平均は一様分布に比べ正規性を示しにくいことがわかる。また、どちらの分布型でも標本サイズとの関係は明瞭ではなかった。

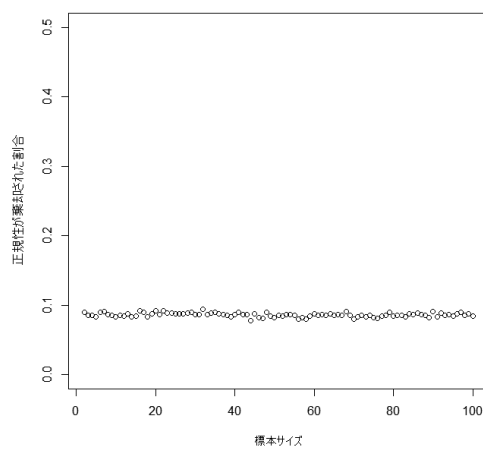


図-4 指数分布から得られた標本平均値の正規性と標本サイズの関係

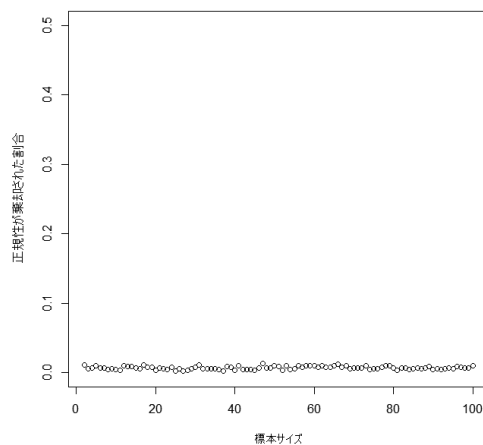
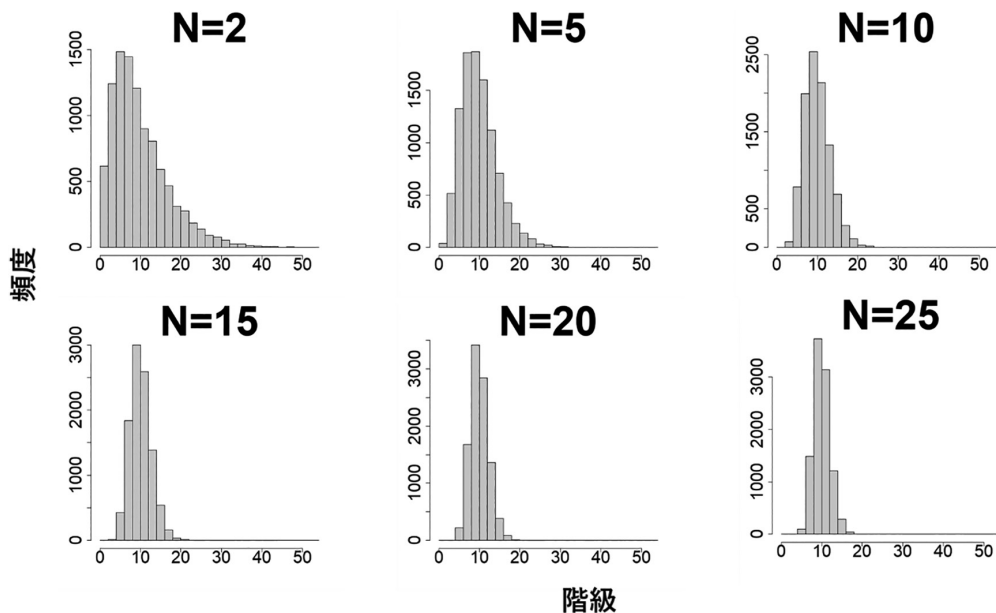


図-5 一様乱数から得られた標本平均値の正規性と標本サイズの関係

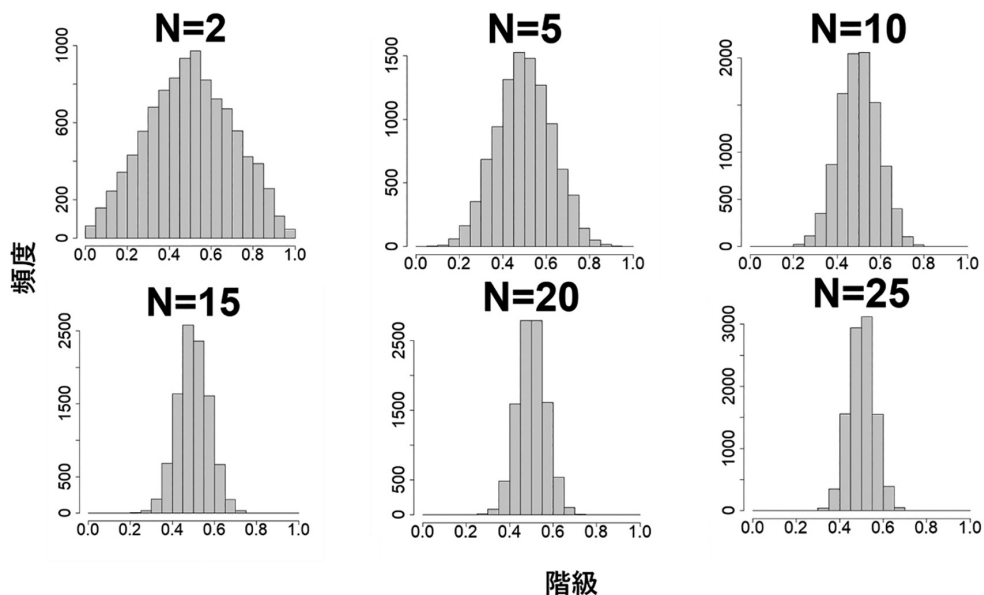
4.3 頻度分布図を描き視覚的に確認する方法

一般的に、このような統計的仮説検定は、帰無仮説の棄却が積極的だが、採択は消極的である（森元，2022）。正規性が棄却されないことが、正規分布を示している事の証拠とはなり得ないと考える。そこで、本節では指数乱数、一様乱数の平均値を標本サイズ毎に 10000 回発生させ、頻度分布図を描き、視覚的に正規性を確認してみた。

非正規分布として指数関数と一様分布を考えた。指数分布は $\text{rate}=1/10$ の指数関数から N 個の変数をサンプリングし、その平均値をとることを 10000 回繰り返し、頻度分布図を作成してみた。また、一様分布は $\text{min}=5$ 、 $\text{max}=15$ とした一様関数から N 個の変数をサンプリングして、同様に 10000 回繰り返し平均値をとった。図－6 に指数分布に関して $N=2 \sim 25$ に変化させ描いたヒストグラムを示す。同様に図－7 に一様分布について示す。



図－6 指数分布から得られた乱数の平均値の分布



図－7 一様分布から得られた乱数の平均値の分布

結果を見ると、指数分布の場合は平均値の分布は、 $N = 15$ までは明らかに右側に長く裾を引く歪んだ形であった。 $N=20$ を超えると左右対称の釣り鐘型に近くなってきたが、やや歪んでいるように見える。 $N=25$ までは明瞭な左右対称の正規分布の形には見えない。これは、一様分布からの平均値の分布と比較すると左右の裾の歪みがより明らかである。一方、一様分布についてみると、 $N=2$ では明瞭なピラミッド型であったが、 $N=5$ ですでに釣り鐘型に近い形状を示していた。また、標本サイズを大きくすると分布の幅は狭くなり、バラつきが抑制されているように見える。また、指数分布の場合と異なり、全ての標本サイズで概ね左右対称の分布型を示しているように見えた。

このように、母集団の分布の形状によって、正規性への収束状況は大きく異なるようだった。特に、指数分布からの標本では標本サイズ $N=20$ 程度までは、平均が正規分布に従わないように見えた。前節での議論では、指数分布を母集団とする標本の場合は、標本サイズを 100 としても、棄却率が 0.05 に届かず、有意水準を過小に見積もるようだった。指数分布は、形状母数 $rate$ を唯一のパラメーターとした典型的な L 字型の分布である。したがって、小標本の場合は標本平均値がばらつきやすいと考えられる。指数乱数の平均値の分布が $N=25$ でも多少の歪みが確認できることから、指数分布と一様分布、指数分布と正規分布の組み合わせで棄却率が高くなるのは、指数分布からの標本が正規分布に収束し難いためではないかと

考える。指数母集団からの標本平均がどのように正規分布に収束するかは今後さらに検討が必要だが、いずれにしても、母集団の分布形状によって標本平均の正規分布への収束の様子が異なっており、結果としてt検定に必要な標本サイズも、母集団の分布形状によって異なってくると推測できる。

5. 効果量、検定力から標本サイズを考える

ここまで、t検定が的確に機能を発揮できる標本サイズを中心極限定理の理論を根拠に検討してきたが、t検定に必要な標本数は母集団分布の形状に大きく影響を受けていることが明らかになった。したがって、検定前に標本の正規性を調べることは意義のあることと言えるが、Shapiro-Wilk 検定等の正規性の検定を行うことは、森元（2022）も指摘しているように対立仮説の採択を意味するものではないので、正規性を保証するものではない。検定の多重性の問題などもあり、ヒストグラムなどの分布から視覚的に正規性について検討するのが最も無難であろう。

一方、正規性が担保されたとしても、有意水準を適正にクリアできるデータ数以上の標本サイズであればよいわけではない。前述のとおり、t検定の性能を左右する要因として、効果量、検定力、標本サイズ、有意水準が挙げられる。検定力は0.8、有意水準は0.05と慣例的に決められている。ここでは、効果量と標本サイズの関係について言及していきたい。

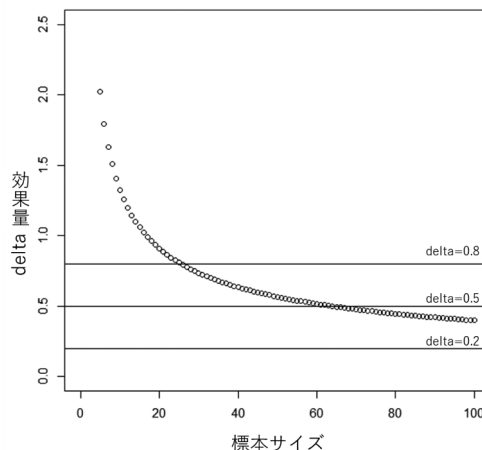
実務において独立した2標本の平均値の差を検証しようとした場合、その差が完全に0であることはあり得ないと考えられる。しかし、以下のt検定量の算出式からもわかる通り、平均の差がわずかな値であっても、標本数を増やすことで検定量は大きくなり、結果p値が小さくなり、帰無仮説が棄却されることになる。

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} = \frac{\bar{x} - \mu}{s} \times \sqrt{n}$$

すなわち、t検定量は、「標準偏差で標準化した平均値の差×標本サイズの平方根」ということになる。つまり、どのようなわずかな差であっても、標本サイズをひたすら大きくとれば、t検定などの検定統計量は必ず有意差を示すのである。つまり、実務上意味のない差であっても、標本サイズを大きくすれば帰無仮説を棄却してしまうということになる。ここから、t検定に関して帰無仮説が棄却されるか否かといったことだけを分析の焦点とするのは意味をなさないとする意見もある（水本ら, 2008, 2010）。そもそも、t検定の目的は「意味のある差が生じているか否か」「2つの標本の母集団は異なる性質を持っているか否か」を明らかにすることであり、差が有るか否かというよりも、どの程度の大きさの差が生じているのか

を客観的に明らかにすることに意味があると考えられる。効果量とは t 統計量を分解した先の式のうち 1 項目の $(\bar{x} - \mu) / s$ であり、標本サイズに依存しない。標本標準偏差で標準化した値であり、2 群にどの程度の差が生じているかを表していると言えるだろう。近年、特に心理学の領域では、独立した 2 標本の差に関して言及する際に検定統計量とともに、標本数に依存しない効果量の検討が必須であるとされてきている（鈴木ら、2012）。

検定力 0.8、有意水準 0.05 としたとき、標本サイズを 5 ～ 100 の範囲で変化させたとき、有意差を検出できる効果量をそれぞれ計算しプロットしたのが図－8 である。計算に用いたスクリプトは参考 4 に示す。



図－8 標本サイズ毎の有意差を検出できる効果量の大きさ

この結果をみると、標本サイズが小さくなるにつれ、検出できる効果量の大きさは大きくなっている。Cohen の目安でいうところの大きな効果量である 0.8 程度の差であれば、30 前後の標本サイズが必要ということになる。すなわち、標本サイズ 30 前後の場合は、データの持つ標準偏差の 0.8 倍程度の差が無いと、帰無仮説を棄却できないということになる。様々な解説書で必要な標本数の目安として 20 ～ 30 程度としているのも、効果量の検討からも整合性が取れると考える。しかし、前述のとおり、母集団分布の形状により、検定の性能が大きく影響を受ける可能性がある。上記のシミュレーションはあくまでも母集団分布が正規分布であることを前提にしている。母集団分布の形状が効果量にどのように影響を与えるかは今後さらなる検討が必要であろう。

6. 総合考察とまとめ～ t 検定における標本サイズをどのように考えればよいか

6.1 中心極限定理から考える標本サイズ

一般的に中心極限定理は、試行回数を増やすと、その値は真の値に近づき、したがって十分な大きさの標本を調べることで、母集団の特徴を正確に把握できると考えられている (Walker, 1969)。ここから、十分な大きさの確率変数の平均値は真値に収束するであろうことは容易に想像つき、その平均値の分布は真値を平均とした正規分布に従うであろうことも納得できる。もし、標本数が十分でなかった場合について考えてみても、平均値は真値周辺に散らばることは疑いの余地がないだろう。しかし、標本数が少ない場合は、真値からのバラつきが大きくなることが推定できる。特に指数分布のような極端に偏った母集団分布の場合は極端にバラつくことが今回のシミュレーションからも推察できる。もちろんであるが、標本数が十分ではないと正規分布と比べ裾の重い分布となってしまう。このようなことから、標本平均が正規分布に従うか否かは標本サイズだけでなく、母集団分布の形状にも大きく影響を受けていることが改めて確認できた。したがって、母集団分布の形状に影響を受けるのであれば、t 検定を適正に行える標本サイズを一概に提示することはできないということになる。

また、母集団分布が正規分布同士、指数分布同士、一様分布同士といった同じ母集団分布からの小標本の場合、棄却率が 0.05 を下回っていたことや、正規分布×一様分布、正規分布×指数分布、一様分布×指数分布といったように異なる形状の母集団からの小標本の場合は逆に棄却率が 0.05 を上回っていたことは注目すべきである。設定した棄却率で安定的に棄却できていないことは明らかであり、p 値が標本サイズと母集団の形状に影響を受けていることの証左であるとも考える。

これらのことから、有意差を正確に検出したいのであれば、母集団の分布形状は予め確認しておくことと同時に、ある程度の標本サイズが必要であろう。概ね標本サイズ 20 未満の場合は棄却率が不安定となり、検定としての精度は高くないと考えられる。厳密な棄却率を保証するものではないが、母集団分布の形状が明らかに異なる場合を除き、少なくとも 20 前後の標本サイズが必要と言えよう。やはり中原 (2022) や多くの統計学の解説書で指摘しているように、やむを得ず小標本で分析しなくてはいけない場合は Wilcoxon の順位和検定や Mann-Whitney の U 検定などノンパラメトリック検定を試すべきであろう。今回検討している t 検定は母集団に特定の分布を仮定して検討するパラメトリック検定の代表的なものだが、ノンパラメトリック検定であれば、検定統計量は母集団の形状に依存しないので、比較的小サイズの標本でも検定可能である。Wilcoxon の順位和検定や Mann-Whitney の U 検定であれば、例えば標本サイズが 4 といった小標本であっても検定可能である。しかし、t 検定のようなパラメトリック検定に比べ、ノンパラメトリック検定は検定力が低くなり、有意な差を見逃

すリスクもある。今後さらに検討が必要であろう。

6.2 効果量から考える標本サイズ

そもそも統計的仮説検定は前述のとおり、標本サイズを大きくすれば検定力が大きくなり、必ず有意差を見出すことができてしまう。したがって、議論の根拠として有意差の有無だけの論文は様々な学会で受理しない傾向にある（水本ら，2008）。有意差の有無ではなく、差の持つ意味が明確になってこそ、分析が有意義となると考える。

また、統計的仮説検定は、分析者自身が持つ仮説（2群の平均値には差がある）を否定する帰無仮説（2群の平均値は等しい）の下で実際の測定値が得られる確率を評価するものである。あくまでも平均値に差が無いと仮定すると「得られた結果は稀な事象である」ということから帰無仮説（2群の平均値は等しい）を棄却するものである。帰無仮説が棄却されたことがすなわち対立仮説を採択することになるかどうかは議論がわかれるところで、分析者としては慎重な態度が求められる（森元，2022）。

このようなことから、標本サイズの決定には、むしろ「求めたい効果量」を分析者自身が設定しておく必要があると考えられる。少なくとも標準化した際の平均値の差が0.8以上あるか否かを検証したいのであれば、少なくとも標本サイズは30以上が必要ということになろう。つまり、適切な標本サイズは一概に設定することはできず、分析者自身が見極めたいと考えている効果量に依存するということである。

6.3 まとめ

以上をまとめると、t検定においては、①母集団の分布を確認すること②どの程度の差を検出したいのか効果量を予め検討しておくこと③母集団分布の形状によって異なる可能性もあり、分布の組み合わせによっては60程度の標本が必要になるケースも一部確認されたが、実務上は極端な母集団分布の形状が異なるケースも稀とも考えられる。分布型が同じ母集団からの標本ということが明らかであれば最低でも20程度、効果量の観点からも考えると30程度の標本サイズは必要と考える④標本サイズが20未満といったサイズになる場合はノンパラメトリック検定を実行したほうが良い といった4点について分析にあたって検討しておくべきと考えられた。

引用文献

Bishop, C.M. (元田 浩・栗田多喜夫・樋口知之ら訳)：パターン認識と機械学習上 ベイズ理論による統計的予測、丸善出版、2012

平松一彦・岡村 寛：小標本における標準誤差と信頼区間の推定、水研センター研報 第6号、

2003、pp.1-8

松原 望・縄田和満・中井検裕：統計学入門、東京大学出版会、1991

水本 篤・竹内 理：研究論文における効果量の報告のために－基礎的概念と注意点－、英語教育研究 31、2008、pp.57-66

水本 篤・竹内 理：効果量と検定力分析入門－統計的検定を正しく使うために－、外国語教育研究のための方法 外国語教育メディア学会 関西支部 メソドロロジー研究部会 2010 年度報告論集、2010、pp.47-73

森元良太：統計的仮説検定の歴史的源流と哲学的意義に関する研究、科学研究費助成事業研究成果報告、2022

中原 治：基礎から学ぶ統計学、羊土社、2022

西内 啓：統計学が最強の学問である [実践編]、ダイヤモンド社、2014

Student : The Probable Error of a Mean., Biometrika Vol. 6, No. 1、1908、pp.1-25

鈴木由美・豊田秀樹：“心理学研究”における効果量・検定力・必要標本数の展望的事例分析、心理学研究第 83 巻第 1 号、2012、pp.51-63

Walker, A.M. : On the asymptotic behaviour of posterior distribution., Journal of the Royal Statistical Society B31(1)、1969、pp.80-88

参考

1 標本標準偏差ヒストグラムの計算

```
simsd <- sapply (1:10000, function(i){ # 定義した関数を 10000 回繰返し simsd に格納
x <- rnorm (3, mean=0, sd=1)          # サイズ 3、平均 0、標準偏差 1 の正規乱数
sd(x)})                               # 発生させた乱数の標準偏差
```

2 rate パラメーターの等しい指数母集団からの標本

```
tsl <- numeric() # 棄却率の格納場所
for (j in 2:100){ # 標本サイズを 2 ～ 100 まで変化させる
alpha <- 0.05    # 有意水準を 5% に設定
m <- j           # 標本 A のサイズ
n <- j           # 標本 B のサイズ
N <- 10000       # 乱数発生回数
```

```

n.reject <- 0          # 棄却回数
for (i in 1:N)         # 試行を N 回繰り返す
{
  x <- rexp(m, rate = 1/10)      # 標本 A として指数乱数を発生
  y <- rexp(n, rate = 1/10)      # 標本 B として指数乱数を発生
  t.stat <- t.test(x, y, var.equal=F, paired=F) # t 検定を実行
  if (t.stat$p.value < 0.05)     # 有意水準を 0.05 として棄却
    n.reject <- n.reject + 1     # 棄却回数をカウント
}
true.sig.level <- n.reject/N      # 棄却率を計算
tsl <- append(tsl, true.sig.level) # 棄却率を tsl に格納
}

```

3 指数乱数の平均値に対して Shapiro-Wilk 検定を実行し p 値を取り出す

```

ratio <- numeric()      # 棄却率の格納場所
for (n in 2:100){
  P <- numeric()         # p-value 格納場所
  for (m in 2:100){      # 乱数発生個数を 2 個から 100 個まで変化させる
    dat <- numeric()     # 乱数平均値の格納場所
    for (i in 1:1000){    # 乱数平均値算出を 1000 回試行
      d1 <- rexp(m, rate=1/10) # rate=1/10 の指数乱数を m 個発生させる
      d2 <- mean(d1)
      dat <- append(dat, d2)}
    s <- shapiro.test(dat) # Shapiro-Wilk 検定の実行
    sp <- s$p.value        # p 値を取り出す
    P <- append(P, sp)}
  ra <- sum(P <= 0.05)/1000 # p 値が 0.05 以下の回数をカウントし割合を計算
  ratio <- append(ratio, ra) # 棄却率を ratio に格納
}

```

4 Cohen の d による効果量と標本サイズの関係

```
d <- numeric()
for (i in 5:100)                # 標本サイズを 5 から 100 まで変化
{
  d1 <- power.t.test(n = i, delta = NULL, power = 0.8, sig.level = 0.05, type = "two.sample",
  alternative = "two.sided")      # Cohen の d は delta に格納される
  d2 <- d1$delta                 # delta を取り出す
  d <- append(d, d2)             # d に格納
}

# 標本サイズとの関係を以下で作図
x <- seq(5,100,1)                # 標本サイズを 5 から 100 で作成
plot(x, d, xlim=c(0,100), ylim=c(0,2.5)) # 散布図を作成
par(new=T)
abline(h=0.2, col=1, lty=1, lwd=1)   # d=0.2 でライン
par(new=T)
abline(h=0.5, col=1, lty=1, lwd=1)   # d=0.5 でライン
par(new=T)
abline(h=0.8, col=1, lty=1, lwd=1)   # d=0.8 でライン
```