

会計実証分析のための検定手法

福井義高

青山学院大学大学院国際マネジメント研究科

〒150-8366 東京都渋谷区渋谷 4-4-25

fukui@gsim.aoyama.ac.jp

2012 年 9 月 14 日

2015 年 11 月 24 日改訂

要旨

係数が厳密にゼロであるという帰無仮説を中心においた会計実証分析の問題点を指摘し、生物学的同等性の考え方に基づき医学・薬学の分野で用いられている検定手法の会計への応用を試みる。

謝辞

統計理論に関して貴重なご助言をいただいた細谷雄三教授及び主催される研究会で発表の機会を与えていただいた深井忠教授及び大日方隆教授にこの場を借りて感謝申し上げます。なお、本稿作成にあたっては科学研究費（20653026；15K13058）の助成を受けた。

Statistical reasoning is an art and so demands both mathematical knowledge and informed judgment.

Gerd Gigerenzer

1. 無益で不確かな仮説検定？

会計研究の主流となりつつある実証研究が依拠する仮説検定は、おそらく多くの研究者が考えているほど理論的に盤石ではない。実際、データ数が増えた今日、以前にもましてその有効性への疑義が多く統計学者から提起されている。たとえば、二つのグループの母平均の差の検定を例に、豊田（2001, 24-25 頁）は以下のように述べる。

帰無仮説の棄却は、危険率を覚悟で仮説を実質的に捨てることであるのに対して、帰無仮説の採択は実質的に態度を留保することである。このような帰無仮説の棄却と採択が対等でないという性質は不合理である。加えて、常識的に考えても、実世界の異なる母数が互いに厳密に等しいということはありません。

データの数が増すと、それに伴って「検定力」と呼ばれる確率がいくらかでも 1 に近づいていく。検定力とは、帰無仮説が偽であるときにそれを棄却する確率である。帰無仮説は先に述べたように、元々、厳密には偽だから、データの数が増せば必ず棄却される。

数理統計学の理論体系の中で大きなウェイトを占めている検定論は、データの洪水に対して、あらゆる意味で無力である。

「実世界の異なる母数が互いに厳密に等しい」というところを、「回帰係数が厳密にゼロに等しい」と読み替えれば、標準的な会計実証研究にもそのまま当てはまる。

ただし、そもそも統計分析の主たる役割は検定ではなく推計なので、会計実証分析もモデルつまり回帰係数の推計に専念すればよいという考え方も一考に値する。検定論の不正確な理解に起因する実証分析の問題点を明快にまとめた Hubbard and Bayarri (2003) は、討論者への返答で次のように指摘している（182 頁）。

We agree that many problems posed and solved as hypothesis testing problems are in reality estimation problems. Only “real” hypotheses (those having some clear, distinctive, meaning and plausibility) should ever enter a hypothesis testing problem. We should never use hypothesis testing tools for a problem that is intrinsically an estimation problem.

とはいえ、会計報告においてある情報を開示するか否かという政策判断は、(ネイマン・ピアソン流の) 意思決定手法としての検定論に比較的親和性があることも否定できない。

そこで、ここでは薬効に関して日本のみならず世界標準となっている「生物学的同等性」(bioequivalence) の考え方¹を参考に、会計実証研究にふさわしい、ひとつの枠組みを提示したい。

ここでは、ネイマン－ピアソン仮説検定の枠組みに厳密に従っているかどうかは怪しいものの、回帰分析を用いてある変数の係数が統計的に有意か否かでその情報価値の有無を判断する、標準的な価値関連性研究そのものの是非は問わない。包括利益と純利益の情報価値の多寡とその開示の必要性をめぐる論争など、会計基準をめぐる実務の動向にも大きな影響を与えているようにみえるこうした実証研究を、現在の研究スタイルを前提に、より意味のあるものすることが本稿の目的である²。

2. 会計実証研究でも明白な二つの問題点

昨今行われている会計実証分析を図式的に表現すれば次のようになる。

株価等の「被説明」変数を y 、「コントロール」変数を x (複数の場合はベクトル)、情報価値の有無が知りたい会計数値を z として、回帰モデル

$$y = \alpha + \beta x + \delta z + \varepsilon$$

を用いて係数を推計し、 $\delta = 0$ の帰無仮説を検定する。モデルの誤差項 ε の分散 σ^2 を一定と仮定すれば、最小二乗推計法により係数の最小分散線形不偏推計値を得ることができる。

こうして得られた係数 δ の推計値 d を、その標準誤差 s_d (ε の推計値であるモデルの標準誤差 s とは別) で割ったもの

$$t = \frac{d}{s_d}$$

が δ の t 値である。漸近理論³に基づけば、帰無仮説が正しい、つまり δ の真の値がゼロの場合、 t は平均ゼロで分散 1 の標準正規分布に従う。推計で得られた t 値をちょうど棄却する検定のサイズが p 値である。

この p 値と事前に設定した有意水準 (危険率) α を比較し、 $p \leq \alpha$ であれば、 $\delta = 0$ の帰無仮説を棄却し、 z には情報価値があると判断し、仮説が棄却できなければ情報価値があるとはいえないと判断する。

こうした会計実証研究手法には、帰無仮説棄却・受容の非対称性とデータ数増加による棄却の必然性という、冒頭に引用した検定論の二つの問題点がまさに「教科書的」に表れている。

¹ 厚生労働省 (2012)。Committee for Medical Products for Human Use (2010)には EU の公式見解が表明されている。

² 本稿の内容は福井 (2014) で一部紹介した。

³ 誤差項に正規分布を仮定し、小標本理論に基づき t 分布を用いても、標本数が極端に少ない限り、 t 分布でも正規分布でも結論はほとんど同じである。

(1) 帰無仮説棄却・受容の非対称性

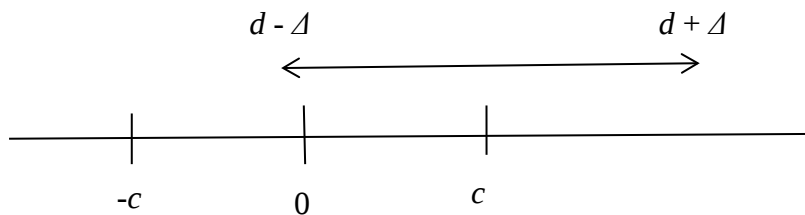
今日、会計研究者がある会計変数の情報価値に関して実証分析を行う場合、研究を進めるうえで念頭にある仮説は、基本的に二つに分けられる。情報価値があるという仮説と情報価値がないという仮説である。

研究者 A は、変数 z には情報価値があると考え、自らの主張とは逆の $\delta = 0$ という帰無仮説が棄却されることを示すことで、自らの主張（変数 z には情報価値がある）を裏付けようとする。実証分析の結果、ある（5 パーセント等、通常小さな）有意水準の下、 $\delta = 0$ という帰無仮説を棄却できたとしよう。この場合、あえて自分に不利な条件を設定したうえで、自らの主張のもっともらしさを示したことになる。

一方、研究者 B は、変数 z には情報価値がないと考え、自らの主張である $\delta = 0$ という帰無仮説が棄却できないことを示すことで、自らの主張（変数 z には情報価値がない）を裏付けようとする。実証分析の結果、ある（5 パーセント等、通常小さな）有意水準の下、 $\delta = 0$ という帰無仮説が棄却できなかったとしよう。この場合、情報価値があるという自らとは逆の主張の証拠が十分でなかったことを示したことになる。

上記の結果から A と B のどちらの主張の説得力が増したであろうか。A は自らに課した高いハードルを越えたと「一応」⁴いえるのに対し、B は相手に課したハードルを相手が越えられなかったことを示しただけである。明らかに、A の主張の方がデータによって説得力を増したといえる。

図 1：推計値 d は「大きい」のに $\delta = 0$ の帰無仮説が棄却できない場合



たとえば、推計値 d の信頼区間 $[d - \Delta, d + \Delta]$ が、かなり「大きな」⁵数値である c とともに

⁴ なぜ「一応」なのかは 2(2)節で明らかになる。

⁵ ここでは漠然と経済的実質からいって無視できない数値と考えていただきたい。また、信頼区間の範囲をどう設定するかという問題は、ひとまず棚上げしておく。

0 を含んでいる図 1 のような場合、 $\delta = 0$ の帰無仮説は棄却できない。とはいえ、変数がある程度の情報価値を含む可能性があることも、通常の言葉の意味では否定できない。

標本数が少ない場合、推計の精度が低いため、信頼区間が広くなることで、ゼロも「大きな」数値も両方、「広い」信頼区間に含まれることは決して珍しくない。むしろこうした実証結果を「情報価値がない」証拠と扱うとすれば、まさに、「帰無仮説の棄却と採択が対等でないという性質は不合理」であることを無視することになる。

そもそも、ネイマン–ピアソン仮説検定においては、棄却あるいは受容がもたらすコストへの考慮が不可欠であり、その観点からも、 $\delta = 0$ の帰無仮説の棄却と受容を「対等」に扱う会計実証研究には問題がある。

会計情報の有用性の有無は、会計基準設定機関がその情報を開示することを求めるか否かを判断する有力な証拠とされている。その場合、有用な情報を非開示とするコストと有用でない情報を開示するコストの比較がなおざりにされていないだろうか。

情報生産コストの劇的低下を考えれば、ある情報の追加的開示コストは無視できる大きさであり、情報開示は投資家にとって有用か否かが判断基準であるべきである。一方で、情報生産コストが低下しても、人間の計算能力には限りがあるため、ある程度経営者が元データを集約して投資家に情報を提供することが、少なくとも一般論としては望ましいので、何でもかんでも情報開示することがよいとは限らない。

しかしながら、投資家の立場からみて有用でなければ、開示された有用でない情報は無視すればよいのに対し、投資家には実は有用であるのに、その情報が非開示となると、投資家が独自に入手することは基本的にできない。開示情報から推計することはできても完全に再現することは通常できないので、経営者が開示すれば避けられたであろう情報ロスが生じる。

したがって、有用でない情報を開示する「間違い」よりも、有用な情報を開示しない「間違い」のコストの方が投資家にとって大きいといってよい。とすれば、図 1 の場合のように、変数がかかりの情報価値を含む可能性があっても、 $\delta = 0$ の帰無仮説が棄却できないことをもって情報価値なしとみなすような実証方法の非合理性は明らかであろう。

(2) データ数増加による棄却の必然性

上記の議論に対して、次のような疑問が生じる。会計実証研究の定番である $\delta = 0$ の帰無仮説による検定では、棄却により有用という判断はできても、棄却できない場合は判断を保留するだけで、有用でないという判断は永遠に出てこない。それでは、少なくとも情報有用性判断基準としては、仮説検定は無意味ということではないのか。

確かに、 $\delta = 0$ の帰無仮説を前提とした仮説検定は、情報価値がないという判断に慎重でなければならないという立場とは相いれない。しかし、問題はそれだけに限られない。データ数が豊富な場合には、 $\delta = 0$ の帰無仮説の検定はさらなる重大な問題を抱えている。棄却により有用と判断することにも必ずしも合理性はないのである。

ここでは議論の本筋を明確にするため、「コントロール」変数のない回帰モデル

$$y = \alpha + \delta z + \varepsilon$$

を用いて、帰無仮説の棄却を情報価値ありの証拠とすることの問題点を明らかにする。

「説明」変数が検定対象の z だけからなる単回帰なので、係数 δ の推計値 d は、

$$d = \frac{\sum (z - \bar{z})y}{\sum (z - \bar{z})^2}$$

その分散と標準誤差はそれぞれ、

$$s_d^2 = \frac{s^2}{\sum (z - \bar{z})^2} \quad s_d = \frac{s}{\sqrt{\sum (z - \bar{z})^2}}$$

となる。

係数の推計値 d やモデルの分散推計値 s の大きさは基本的に標本数に依存しない⁶けれども、

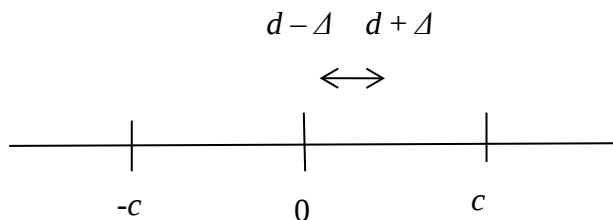
$$\sum (z - \bar{z})^2$$

は大ざっぱにいて標本数 N に比例する。つまり、 t 値

$$t = \frac{d}{s_d} = \frac{d}{\frac{s}{\sqrt{\sum (z - \bar{z})^2}}} = \frac{d}{s} \times \sqrt{\sum (z - \bar{z})^2}$$

は、標本数の平方根 $\sqrt{N}$ に比例する。会計実証研究において十分な考慮が払われていないものの、仮説検定に用いる t 値は標本数の増加に対して中立ではないのだ。

図2：推計値 d は「小さい」のに $\delta = 0$ の帰無仮説が棄却される場合



⁶ 標本数が少ないと推計値が安定しないことは別の問題である。

したがって、 t 値の分子である係数の推計値が極めて小さくても、データが増えると分母である係数の標準誤差がいくらでも小さくなるので、通常使われるような有意水準（たとえば、1 パーセントでも）はハードルとして低すぎるのである⁷。図 2 で示したように、標本数が多くなると信頼区間は狭くなるので、係数が「小さく」ても、厳密にゼロであるという帰無仮説は棄却される。

係数が正確にゼロに等しいという $\delta = 0$ の帰無仮説は、冒頭に引用したとおり、「元々、厳密には偽だから、データの数が増せば必ず棄却される」。

会計データには、検定論が前提とする標本抽出の誤差とは別に、測定に起因する非標本誤差が含まれる。したがって、林（1984, 129-130 頁）が指摘するように、「標本数が十分少ない時には、標本抽出誤差が十分大きく非標本誤差のほうが比較的に小さいとみなされるし、実質的に大きな差がないと有意とならないので役立つことが多い」ものの、標本数が大きくなると、「標本抽出による計算された分散は十分小さくなり、このくらいの差は実質的な差といえないという限界を下まわり、また大概のものは非標本誤差がずっと大きくなるので、検定論の問題は枝葉末節—むしろ無意味というより非科学的—なことになる」。

また、Hansen（2015, 192 頁）は経済学博士課程用の計量経済学テキストで、会計実証研究者にとって耳の痛い忠告を行っている。

In many modern applications, sample sizes are very large so standard errors can be very small. Consequently t -ratios can be large even if the coefficient estimates are economically small. In such contexts it may not be interesting to announce “The coefficient is non-zero!” Instead, what is interesting to announce is that “The coefficient estimate is economically interesting!”

In particular, some applied papers report coefficient estimates and t -ratios, and limit their discussion of the results to describing which variables are “significant” (meaning that their t -ratios exceed 2) and the signs of the coefficient estimates. This is very poor empirical work, and should be studiously avoided. It is also a recipe for banishment of your work to lower tier economics journals.

3. 厳密にではなく実質的に効果があるか否か

データ数が少なければ、棄却と受容の非対称性から「効果なし」という結論の脆弱さが明らかであり、逆にデータ数が多ければ、棄却という答えが最初からわかっている。係数が厳密にゼロとする帰無仮説を中心に組み立てられた、会計実証分析で用いられる分析手法には、根本的欠陥があるといつてよい。そこで、薬効に関して用いられる生物学的同等性を参考に、統計的検定論の枠組みの中で、新たな会計実証分析手法を提示したい。

⁷ Arrow (1960) はハードルつまり有意水準を標本数に合わせて増やすという解決策を提示している。

厚生労働省のガイドライン⁸によれば、「生物学的同等性試験では、通常、先発医薬品と後発医薬品のバイオアベイラビリティを比較する。それが困難な場合、又は、バイオアベイラビリティの測定が治療効果の指標とならない医薬品では、原則として、先発医薬品と後発医薬品との間で、効力を裏付ける薬理作用、又は、主要効能に対する治療効果を比較する」。要するに、先発医薬品と後発（ジェネリック）医薬品の薬効が同じかどうかを試験するわけである。差がないと認められれば、後発薬は市販することが認められる

この場合、薬のいろいろな特性が両者で同じかどうかを、両者の差が厳密にゼロかどうかという帰無仮説で判断すると、後発薬が認可される可能性は事実上、閉ざされる。なぜなら、2(2)節で述べたように、標本数が多ければ、帰無仮説の棄却つまり「差がある」という結果が「約束」されているからだ。

もちろん、薬効が同じなのに異なると判断して後発薬発売を不許可とするより、薬効が異なるのに同じと間違えて判断して後発薬の発売を認めるほうがコストあるいはリスクは大きい。前者の場合は、すでに先行薬が存在するので、高価であったとしても、後発薬がなくても、患者の治療に悪影響をもたらすことはないのに対し、後者の場合は、薬効が異なる後発薬が使われることで、治療に悪影響をもたらす得る。

一方で、医学・薬学上、薬の各種特性値が厳密には同じでなくとも、実質的には同じと行ってよい幅がある。差がこの幅に入っていれば同じと判断することとすれば、標本数が増えることによって推計の精度が増し、合理的な意思決定を行うことができる。

そこで、「薬効が違う」という仮説を帰無仮説として、この帰無仮説が低い有意水準で棄却された場合のみ、後発薬発売を認めることとすれば、薬効が異なる後発薬が市場に出回るというリスクを直接コントロールできる⁹。

これと同じ考えは会計基準設定機関が会計情報の有用性を判断する際にも応用できる。

我々が知りたいのは、測定における誤差も考慮したうえで、会計データに情報価値があるか否か、つまり係数 δ の値が会計上、実質的に意味のある大きさかどうかであって、 δ が正確にゼロに等しいかどうかということではない。ここまでの議論で示したように、 $\delta = 0$ という単純帰無仮説の検定では、我々の知りたいことは明らかにできない。

むしろ、 δ の絶対値（したがって正の値）が一定の数値 c を超える、つまり $c \leq \delta$ か $\delta \leq -c$ のどちらかが満たされるか否かで、会計データの実質的情報価値の有無を判断するほうが理にかなっている。

ここでは、こうした考え方を伝統的検定論の枠組みの下で定式化する前に、会計情報の有用性判断を完全に統計的問題として定式化することはできないことを指摘しておきたい。ここで展開する「会計的有用性」とでもいうべき考え方は、会計上、実質的に意味のある（係数の）大きさはどれくらいかという、統計理論だけでは決められない、会計固有の判

⁸ 厚生労働省（2012, 4 頁）。

⁹ より具体的に表現すれば、実質的に影響を及ぼす係数下限（の絶対）値を c とすれば、 c より大きいという帰無仮説が棄却された場合のみ、後発薬は承認される。

断が必要となる。昨今、高度な統計手法を使い、その数値の大小を問わず、係数が統計的に有意かどうかで、機械的に情報価値がある（あるいはない）という結論を下すのが、会計実証研究の定番のようである。会計実証研究のアルゴリズム化である。

しかし、数値の大小の意味を議論せず素通りすることは、測定の学問たる会計の研究者としては、一種の自殺行為であろう。こうした批判に対し、「客観的」に決められる統計的有意性と異なり、どのような大きさならば実質的に意味があるのかは主観的にならざるをえず、実証科学を目指す会計研究の潮流とは相いれない、という反論が予想される。

実は、「客観的」にみえる通常の統計的有意性の議論も、有意水準をいくりにするかで主観的にならざるを得ない。さらに、すでに指摘したように、標本数が増えると t 値は大きくなるので、同じ t 値であっても、標本数が異なれば、その持つ意味は違ってくる。この Arrow (1960) が半世紀以上前に提起した問題は、会計実証研究では事実上、無視されたままである。

そもそも、どのような大きさであれば実質的に意味があるのかを議論することこそ、会計研究者の固有の領域だとすらいえる。絶対正しい答えがなくとも、専門家としての知見を活かし、一定の範囲でのコンセンサスを得ることは不可能ではない。有用な情報を排除するリスクにより一層の考慮を払うのが望ましいとすれば、常識的な範囲での最小値を基準（上記の c ）にすればよい。

4. 会計的有用性の定式化¹⁰：二重片側検定

2(1)節で指摘したように、会計データに関しては有用でない情報を開示する「間違い」よりも、有用な情報を開示しない「間違い」の方が投資家にとってコストが大きい。したがって、実質的に効果があるという仮説が積極的に否定（棄却）されてはじめて、情報価値がないと結論付けるべきである。逆にいえば、実質的にゼロであることが否定（棄却）できないだけで情報価値がないと結論付けるべきではない。

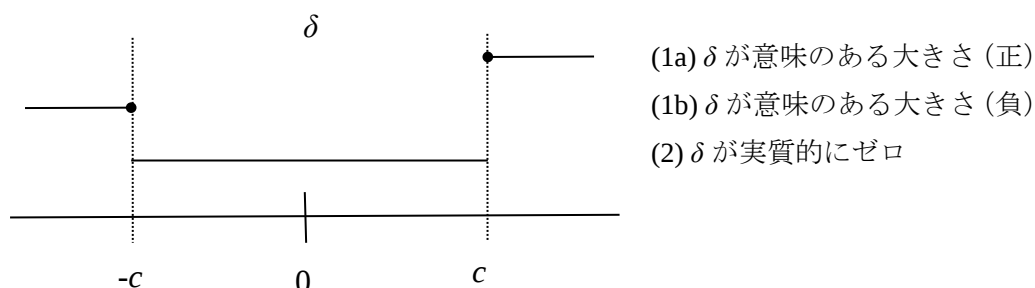
要するに意味のある数値（の絶対値）の下限である c よりも母数 δ （の絶対値）が大きいことはまずありえないことが示された場合、すなわち図 3(1a)の $c \leq \delta$ と (1b)の $\delta \leq -c$ という二つの帰無仮説が両方とも棄却された場合のみ、情報価値がないと結論付けるのである。その場合は、(2)の $-c < \delta < c$ であると結論付けたといってもよい。

係数 δ が意味のある大きさであるという帰無仮説は $\delta \leq -c$ あるいは $c \leq \delta$ のどちらかが成り立てばよいので、棄却するには両方の条件（仮説）を棄却せねばならないことに注意していただきたい。二つの片側検定を同時に行うことから、「二重片側検定」(two one-sided tests, TOST) と呼ばれる¹¹。

¹⁰ 以下で述べる統計手法の解説は、生物学的同等性を念頭に置いた Berger and Hsu (1996) と Shuirmann (1987) に負っている。

¹¹ ここで解説した TOST に比べ理論上、検定力の点ですぐれた手法が存在するものの、Hauk and Anderson (1996) が指摘するように、そうした手法の実際上の優位性はないに等しい。したがって、TOST の簡便性を考慮すれば、それ以外の手法を用いる実利性はないと思われる。

図3：母数 δ が会計的に意味のある大きさ（絶対値が c 以上）か否か



ある標本から母集団の係数を推計する以上、推計誤差は避けられない。したがって、係数推計値（の絶対値）が意味のある数値（の絶対値）の下限より「十分」小さくない限り、帰無仮説は棄却できない。その基準が有意水準 α である。

まず、母数 δ が（正の値である） c より大きいという帰無仮説を棄却するには、推計値 d が c より「十分」小さくしなければならない。有意水準が α の場合、

$$t_U = \frac{c-d}{s_d} \geq t_{1-\alpha}$$

が満たされれば、 $c \leq \delta$ という仮説は棄却される。

次に、母数 δ が（負の値である） $-c$ より小さいという帰無仮説を棄却するには、推計値 d が $-c$ より「十分」大きく（絶対値で考える場合は、小さく）なければならない。有意水準が α の場合、

$$t_L = \frac{d-(-c)}{s_d} = \frac{c+d}{s_d} \geq t_{1-\alpha}$$

が満たされれば、 $\delta \leq -c$ という仮説は棄却される。

Berger (1982)は、 k 個の帰無仮説がそれぞれ有意水準 α で検定されれば、個別の帰無仮説の少なくとも一つが正しいという帰無仮説の有意水準もやはり α となることを示した¹²。したがって、二つの帰無仮説 $c \leq \delta$ および $\delta \leq -c$ がそれぞれ有意水準 α で棄却されれば、係数 δ が会計的に意味のある数値であるという仮説、すなわち $c \leq \delta$ あるいは $\delta \leq -c$ という仮説も有意水準 α で棄却される。

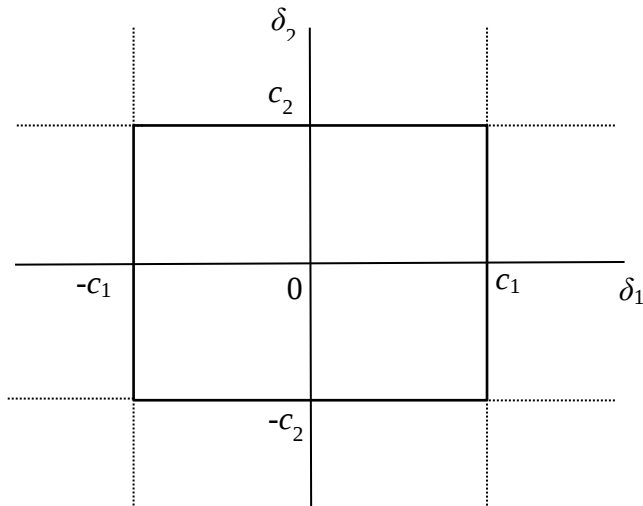
我々が関心を持つ会計変数が複数の場合も同様の議論が可能である。たとえば、ある会

¹² 証明は補論参照。

計基準の下では、二つの変数 z_1 および z_2 がセットで開示されるとしよう。これは $k=4$ の場合に相当する。この会計基準が投資家に有用な情報をもたらすか否かは、二つの変数の係数 δ_1 と δ_2 の少なくともどちらかが会計的意味のある大きさ（絶対値がそれぞれ c_1 と c_2 以上）となるか否かで判断される。

図示すれば、図 4 で (δ_1, δ_2) が 0 を中心とした長方形の外側にあれば、この会計基準は有用と判断される。もし、 $c_1 \leq \delta_1$ 、 $\delta_1 \leq -c_1$ 、 $c_2 \leq \delta_2$ および $\delta_2 \leq -c_2$ という四つの帰無仮説すべてが有意水準 α で棄却された場合、すなわち (δ_1, δ_2) が長方形の内側にありと判断されれば、会計基準に情報価値があるという仮説がやはり有意水準 α で棄却される。変数が 3 個以上の場合も同様であることは自明であろう。

図 4：母数 δ_1, δ_2 が会計的に意味のある大きさ（絶対値が c_1, c_2 以上）か否か



5. 会計的有用性の定式化：100(1-2 α)パーセント信頼区間

上記と同じ検定を、より直感的に理解しやすいかたちで示すことができる。

有意水準における t 値に係数推計値 d の標準誤差 s_d をかけたものを

$$\Delta = t_{1-\alpha} \cdot s_d$$

と定義すれば、

$$t_U = \frac{c-d}{s_d} \geq t_{1-\alpha}, \quad t_L = \frac{d-(-c)}{s_d} = \frac{c+d}{s_d} \geq t_{1-\alpha}$$

という条件は、それぞれ、

$$c-d \geq \Delta, \quad c+d \geq \Delta$$

と表わせる。さらに変形すると、

$$d+\Delta \leq c, \quad -c \leq d-\Delta$$

となる。

情報価値があるという仮説が棄却されるには、両方の条件が満たされなければならない。すなわち、 $[d-\Delta, d+\Delta]$ が会計的にはゼロといってよい範囲¹³である $[-c, c]$ に含まれれば、有意水準 α で仮説は棄却される。

ところで、 Δ は片側検定での有意水準 α に基づいているため、 $[d-\Delta, d+\Delta]$ は d を中心とした、有意水準 α ではなく 2α の信頼区間になっている。

要するに、有意水準 α のTOSTは、 $100(1-2\alpha)$ パーセント信頼区間が $[-c, c]$ に含まれるか否かと同値である。係数に会計的意味があるか否かは、推計値 d を中心とした $100(1-2\alpha)$ パーセント信頼区間と実質的にゼロの範囲を比較すればよい。

もちろん、生物学的同等性の場合同様、どのような大きさであれば会計的意味があるのか、つまり c を設定する必要がある。しかし、こうした仮説検定の進め方を採用しないにせよ、係数 δ の点推定値 d を中心とする信頼区間の解釈に限っても、TOSTの考え方は従来とは違って視点を提供する。

両側検定の発想¹⁴に基づく通常の信頼区間解釈においては、点推定値 d が得られた場合、信頼区間の上下限を片側検定有意水準 $\alpha/2$ に相当する $\Delta_{\alpha/2}$ を用いた、 $[d-\Delta_{\alpha/2}, d+\Delta_{\alpha/2}]$ が有意水準 α の信頼区間とされている。

しかし、係数 δ が（実質的に）ゼロであるという仮説を棄却するのではなく、係数 δ が実質的に意味のある大きさを持つという仮説を棄却するという角度からアプローチするのであれば、第5節で述べたように、有意水準 α の信頼区間は、上下限に片側検定有意水準 α に相当する Δ_α を用いた、 $[d-\Delta_\alpha, d+\Delta_\alpha]$ となる。つまり、通常の場合よりも幅が狭くなる。

信頼区間として、 $[d-\Delta_{\alpha/2}, d+\Delta_{\alpha/2}]$ ではなく、 $[d-\Delta_\alpha, d+\Delta_\alpha]$ を用いるというのが、有意水準 α の検定には $100(1-2\alpha)$ パーセント信頼区間が対応するという意味である。

情報価値がないということを示したい研究者は、情報価値がないという自らの主張が否定できないという低いハードルを設定するのではなく、自らの実証研究のハードルを高くする、つまり情報価値があるという仮説を帰無仮説にとることが望ましい。この観点からは、 $100(1-2\alpha)$ パーセント信頼区間を用いることが望ましいということになる。

6. 常識に合致する仮説検定

TOST（あるいは $100(1-2\alpha)$ パーセント信頼区間）を用いれば、従来の厳密に係数ゼロという帰無仮説を用いた場合の問題点を回避した、常識に合致した検定を行うことができる。

図5と図6は、係数 δ が会計的意味のある大きさか否かという考え方にに基づき、変数 z に情報価値があるという帰無仮説が棄却される場合とされない場合を図示したものである。

¹³ 正確には、意味のある下限値である両端点を含む。

¹⁴ 以下の議論は、本来の仮説検定論の立場とは必ずしも整合しないので、「両側検定的」という表現を用いた。

図 5 : 変数 z に情報価値があるという帰無仮説が TOST で棄却される場合

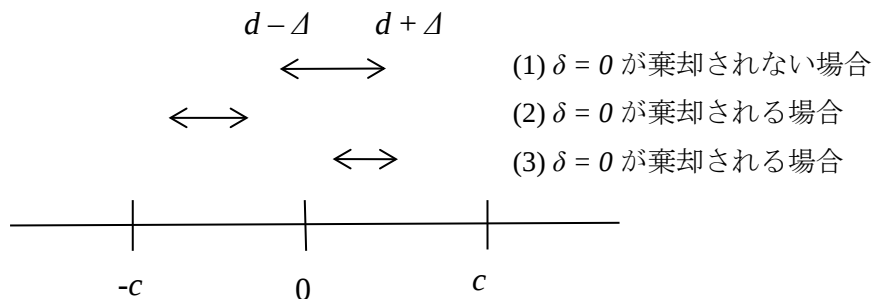
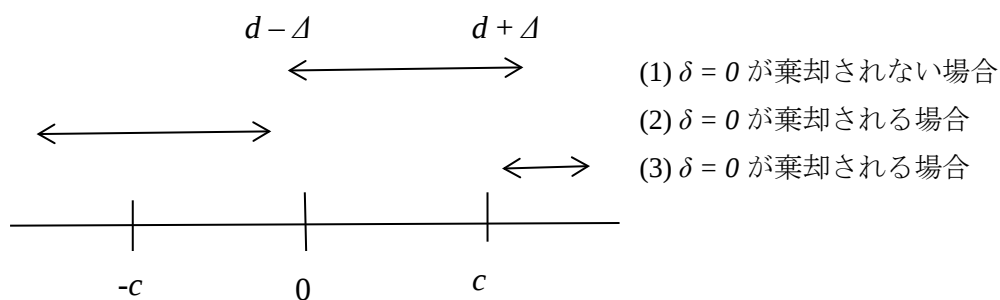


図 6 : 変数 z に情報価値があるという帰無仮説が TOST で棄却されない場合



たとえば、厳密に係数 δ がゼロの帰無仮説が棄却されたことをもって情報価値ありとする通常の検定手法を用いると、図 5(2)および(3)の場合は情報価値あり、図 6(1)の場合は情報価値なしという、TOST とは反対の結論となる。

図 5(2)および(3)の場合は、係数の推計値自体は（常識的な意味で）小さいのに、標準誤差が小さいため、 t 値が名目的に大きくなることで厳密にゼロという仮説が棄却されることになる。「帰無仮説は... 元々、厳密には偽だから、データの数が増せば必ず棄却される」という冒頭の批判が当てはまる場合である。

一方、図 6(1)の場合は、係数の推計値自体は（常識的な意味で）大きい可能性があるのに、係数の標準誤差が大きいので、信頼区間が長くなりゼロも含まれるため、係数がゼロとい

う仮説が棄却できない、つまり係数ゼロという反対仮説が受容される。「帰無仮説の採択は実質的に態度を留保することである。このような帰無仮説の棄却と採択が対等でないという性質は不合理である」という批判が当てはまる場合である。

それに対し、図 5 と図 6 の例を見れば、係数が厳密にゼロという帰無仮説を用いた場合の問題点二つを TOST が解決していることがわかる。TOST は常識に合致する検定手法ということができる。

7. 結論：会計実証研究に向けて

TOST は、実証研究者の間で多用される統計的検定論の問題点を回避するとともに、実証分析によって得られた推計値がどのくらいの大きさならば会計的に意味があるのかという問いかけを通じて、実証研究者に会計学者としての貢献を期待する手法でもある。本稿を読んで、TOST を使うかどうかはともかく、係数が厳密にゼロという帰無仮説を中心とする実証分析が抱える致命的といってよい二つの問題点—帰無仮説棄却・受容の非対称性とデータ数増加による棄却の必然性—に気づいて頂ければ幸いである。

補論：定理 (Berger 1982, Theorem 1) の証明

推計の対象となる未知の母数（スカラーあるいはベクトル） θ がある範囲 Θ_i ($i = \{1, 2, \dots, k\}$)に属するという k 個の帰無仮説

$$H_{0i} : \theta \in \Theta_i$$

を考える。このうち少なくとも一つの帰無仮説が正しいという仮説 H_0 は

$$\theta \in \Theta_0 \equiv \bigcup_{i=1}^k \Theta_i$$

という主張と定式化できる。

さて、範囲（集合） W の補集合を W' と表わそう。仮説 H_0 を帰無仮説とすれば、その対立仮説 H_A は

$$\theta \in \Theta_A \equiv \bigcap_{i=1}^k \Theta'_i$$

となる。

たとえば、2個の帰無仮説を $H_{01} : \theta \leq -c$ と $H_{02} : c \leq \theta$ とすれば、 H_0 はそのどちらか成り立てばよく、その対立仮説 H_A は両方満たされない場合、すなわち $-c < \theta < c$ となる。

仮説 H_{0i} 検定の棄却領域を R_i とすれば、有意水準（危険率） α の検定とは、仮説が正しいのに棄却される確率 P が、すべての $\theta \in \Theta_i$ において、

$$P(R_i | \Theta_i) \leq \alpha$$

となっていることを意味する。

それでは、 k 個の帰無仮説 H_{0i} に有意水準 α の検定を行った場合、この k 個の検定は H_0 の検定として、どれだけの有意水準を確保できるか。

仮説 H_0 検定の棄却領域 R は k 個の検定全てにおいて棄却領域 R_i に含まれていなければならない、つまり

$$R = \bigcap_{i=1}^k R_i$$

となり、当然、 $R \subseteq R_i$ である。一方、仮説 H_0 を満たす母数 $\theta \in \Theta_0$ は、 k 個の帰無仮説 H_{0i} のうち少なくともどれか一つの（その仮説を満たす）母数領域 Θ_i に含まれる。

したがって、個々の帰無仮説 H_{0i} の有意水準が α であれば、すべての $\theta \in \Theta_0$ において、

$$P(R | \theta) \leq P(R_i | \theta) \leq \alpha$$

が成り立つ。すなわち、仮説 H_0 の有意水準は最大でも α となる。

たとえば、2個の帰無仮説 $H_{01} : \theta \leq -c$ と $H_{02} : c \leq \theta$ を、それぞれ有意水準 α で検定すれば、 $\theta \leq -c$ あるいは $c \leq \theta$ という仮説 H_0 も有意水準 α （と同じかそれより小さな水準）で検定されたことになる。なお、 H_{01} と H_{02} が両方とも、つまり H_0 が棄却されれば、対立仮説 H_A 、つまり $-c < \theta < c$ が受容されることになる。

参考文献

- 厚生労働省 (2012) 「後発医薬品の生物学的同等性試験ガイドライン等の一部改正について : (別添) 後発医薬品の生物学的同等性試験ガイドライン」 薬食審査発 0229 第 10 号。
- 豊田秀樹 (2001) 『金鉱を掘り当てる統計学 : データマイニング入門』 講談社。
- 林知己夫 (1984) 『調査の科学 : 社会調査の考え方と方法』 講談社。
- 福井義高 (2014) 「ひょっとすると役に立つかもしれない会計のはなし⑨ : 《使える》 仮説検定」 『企業会計』 66 巻 12 号 76-81 頁。
- Arrow, K. J. 1960. Decision Theory and the Choice of a Level of Significance for the t -Test. In I. Olkin et al. (eds.), *Contribution to Probability and Statistics: Essays in Honor of Harold Hotelling*. Stanford, U.S.A.: Stanford University Press.
- Berger, R. L. 1982. Multiparameter Hypothesis Testing and Acceptance Sampling. *Technometrics* 24 (4): 295-300.
- Berger, R. L., and J. C. Hsu. 1996. Bioequivalence Trials, Intersection-Union Tests and Equivalence Confidence Interval. *Statistical Science* 11 (4): 283-302.
- Committee for Medical Products for Human Use. 2010. *Guideline on the Investigation on Bioequivalence*. European Medicines Agency.
- Hansen, B. E. 2015. Econometrics, Lecture Notes, Department of Economics, University of Wisconsin.
- Hauk, W. W., and S. Anderson. 1996. Comment on Berger and Hsu, Bioequivalence Trials, Intersection-Union Tests and Equivalence Confidence Interval. *Statistical Science* 11 (4): 303.
- Hubbard, R., and M. J. Bayarri. 2003. Confusion over Measures of Evidence (p 's) versus Errors (α 's) in Classical Statistical Testing (with Discussion and Rejoinder). *American Statistician* 57 (3): 171-182.
- Schuurmann, D. J. 1987. A Comparison of the Two One-Sided Tests Procedure and the Power Approach for Assessing the Equivalence of Average Bioavailability. *Journal of Pharmacokinetics and Biopharmaceutics* 15 (6): 657-680.