日本の心理学研究論文における Mann-WhitneyのU検定の誤用とその対策

富　原　一　哉

Mann-WhitneyのU検定使用における前提条件

Mann-WhitneyのU検定（以下U検定）は，独立する2標本間の中央値の差の検定として，最も頻繁に用いられるノンパラメトリック検定の一つである。ノンパラメトリック検定であるので，一般に母集団の分布特性に関わりなく検定を用いることが出来ると解されている。しかしながら，U検定は確かに母集団の分布の型に関しては特定の仮説を持たないものの，他の多くの順位に基づくノンパラメトリック検定と同様，比較に用いる2つの標本集団の元となったそれぞれの母集団における散布度の等質性を前提としている（Kasuya，2001；Siegel & Castellan,1988）。U検定の帰無仮説は，「2つの互いに独立な累積分布関数F（x），G（y）は同じである（岩原，1964）」であり，これは2つの標本が同一の母集団から得られていることと同義である。したがって，散布度においても両者が一致することが当然求められる（e.g. Kasuya，2001；Siegel & Castellan，1988）。

実際，Kasuya（2001）は，2標本間の散布度が等質ではない場合にU検定を用いると，第1種の過誤の確率が上昇することをコンピュータ・シミュレーションによって示している。彼のシミュレーションによれば，両群の分散の比が4であり，標本の大きさが大きく異なる場合（n_1=30，n_2=10）には，第1種の過誤の確率は最大で14％を超えることが示されている。しかしながら，この点の認識は低く，Kasuya（2001）は，「Animal Behaviour」誌56巻（1999）の中で，全論文の約30％にあたる50論文でU検定やその拡張型であるKruskal-WallisのH検定が用いられており，そのうち散布度の等質性について検証できた21論文中15本において散布度が有意に異なっていることを示

した。これは動物行動研究においては散布度が等質ではない場合でもU検定やKruskal-WallisのH検定が頻繁に用いられていることを示唆している。

心理学研究におけるU検定の使用

翻って，もう一つの行動研究の分野である心理学においてはどうであろうか。動物行動学と比較して，心理学においてはU検定の使用はそれほど頻繁ではない。2002年および2003年に発行された「心理学研究（第73巻，74巻）」，「Japanese Psychological Research （第44巻，45巻）」，「動物心理学研究（第52巻，53巻）」の3誌に掲載された原著および資料論文計183本のうち，U検定およびその拡張型であるKruskal-WallisのH検定を用いていたのはわずか5本であった。しかしながら，使用頻度が低いからと言って先の誤用の問題を無視するわけにはいかない。また，U検定の使用頻度の低さは，本来ノンパラメトリック検定が用いられる方が望ましい場合でも，いわば強引にt検定や分散分析などのパラメトリック検定を適用することが多いことがその一因となっていると思われる。したがって，適切にU検定が使用されるようになれば，前述の散布度の等質性の問題はさらに増すものと思われる。

実際，先に上げたU検定およびKruskal-WallisのH検定を使用した5本の論文の中で，それらの検定を使用するにあたって前提条件である散布度の等質性が検定されたかどうかは明確ではない。また，論文では分散等，散布度についての正確な数値が与えられていないので検定は不可能であるが，少なくとも1論文では図中に示されたerror barから散布度の比が4を超えることを予測させる結果が示されていた。したがって，心理学研究においても動物行動研究と同じ誤用の危険性が存在すると言ってよいだろう。

統計テキストにおける取り扱われ方

心理学研究において用いられている統計のテキストを全て調べることは困難であるが，著者の手元にある日本語のテキストのうち，U検定について取り扱っている9冊を調べたところ，U検定の帰無仮説として「2つの母集団

が同じである」ことを明確に説明しているものは7冊であり，残り2冊は単に「中央値が等しい」と述べるか，実際の計算方法のみで明確な帰無仮説を示していなかった（これらのテキストのうち本稿に引用のないものについては，最後に参考文献として列記した）。さらに，U検定の使用にあたってその前提条件として散布度の等質性を確認する必要性を記述しているテキストはなかった。U検定に対応する「2つの独立した標本の平均値の差」についてのパラメトリック検定であるt検定については，ノンパラメトリック法のみを取り扱った2冊を除く全てのテキストで，その前提条件として分散の等質性が記述してあることを考えると，U検定における散布度の等質性の問題は非常に認識が低いと言わざるを得ない。

さらに，岩原（1984，1985）はt検定において分散が等質でない場合の対処法の一つとして「ノンパラメトリック検定を用いる」ことを挙げ，その中でU検定を紹介している。また，類似の記述は最近のテキストの中にも認められる（例えば 森，1999）。特に岩原（1984）の著書は本邦における推測統計学の古典的名著であり，教育・心理学分野のみならず医学・社会学等多くの行動科学分野にその影響を与えた。このような著書において誤用を招くような記述のあることは非常に残念であり，これが心理学においてU検定における散布度の等質性の問題の認識を遅らせる一因となった可能性は否定できないだろう。

散布度の検定と対処法

それでは，U検定において2つの母集団の散布度が等質ではないために起こる第1種の過誤の上昇を防ぐためにはどうすればよいだろうか。Kasuya（2001）は，U検定の前にSiegel-Tukey検定を行い散布度の等質性を確認した上でU検定を用いることを提唱している。これは，独立する2標本の平均値の差をt検定によって検討する場合に，その前段階としてF検定を用いて分散の等質性を確認する作業と同等である。彼のシミュレーションによれば，Siegel-Tukey検定を行った上で散布度に有意差の認められなかった対に

のみU検定を行った場合，第1種の過誤の率が5％を超えることはほとんどなく，母集団の分散の差が大きくなると，むしろ減少する傾向にあることが示された。しかしながら，t検定の場合と同様，実際に散布度に有意差が認められた場合はそのままU検定を用いることが出来ない。したがって，散布度の等質性を前提としないノンパラメトリック検定を代わりに用いる必要がある。Kasuya（2001）はそのような検定手法の一つとして中央値検定（median test）を上げている。Siegel-Tukey 検定も中央値検定も心理学においてはあまり一般的なものではないので，以下にその具体的な検定手順をまとめておく。なお，詳細な手順や検定例については岩原（1979），田中・垂水（1999）を参照のこと。

Siegel-Tukey 検定

Siegel-Tukey 検定は，2組の独立な標本に基づいて，2つの母集団の散布度の差を検定するノンパラメトリック検定の一つである。2つの母集団の分布尺度パラメータをそれぞれσ_1，σ_2とすると，帰無仮説は「H_0: 2つの母集団分布は同じ（$\sigma_1=\sigma_2$）」であり，対立仮説は「H_1: 2つの母集団分布は異なる（$\sigma_1\neq\sigma_2$）」となる。ここで，2組の独立な標本集団A［$X_{11}, X_{12}, \cdots X_{1n_1}$］（$n_1$は標本Aの大きさ）とB［$X_{21}, X_{22}, \cdots X_{2n_2}$］（$n_2$は標本Bの大きさ）が得られたとすると，まずその両方をこみにして，最小値から最大値まで観測値を順に並べる。次に，それらの観測値に対して以下の手順で順位をつける。

(1) 最小値に順位1をつける。
(2) 最大値に順位2をつける。
(3) 2番目に大きな値に3をつける。
(4) 2番目に小さな値に4をつける。
(5) 3番目に小さな値に5をつける。
(6) 以下この要領で，小→大→大→小→小→大…とすべての観測値に順位をつける。
(7) 2つの標本の大きさの合計（$n_1 + n_2$）が奇数の場合は，最後に残る全体

の中央値は順位をつけずに無視する。

次にこうして得られた観測値の順位を再度元のA群とB群に分け，各群の順位和R_1およびR_2を算出する。もし両群の散布度に差がないならば，こうして得られた2群の順位和は等しくなるはずである。そこで，以降はこのR_1，R_2を用いて，U検定と同じ要領で検定する。

中央値検定

中央値検定は2組の独立な標本にもとづいて2つの母集団の中央値が等しいかどうかを検定するノンパラメトリック検定法の一つである。中央値検定では前提条件が特になく（岩原，1979），したがって，2つの母集団の散布度が異なりU検定が用いられない場合にも適用することが出来る。2つの母集団の中央値をそれぞれM_1，M_2とすると，帰無仮説は「2つの母集団の中央値は等しい$(M_1 = M_2)$」であり，対立仮説は「2つの母集団の中央値は異なる$(M_1 \neq M_2)$」となる。ここで，2つの標本集団をこみにして全体の中央値を求め，各群に対して中央値より上と下の度数をそれぞれ求め，2×2の分割表を作成する。もし，両群の中央値が等しければ，ともに中央値を超える値と超えない値の度数は等しく，一致するはずである。そこで，この2×2の分割表に対して，度数の偏りについてχ^2検定を用いて検定する。ただし，2つの標本の大きさの合計（$n_1 + n_2$）が10以下であれば，Fisherの直接確率法を用いる方がよい。

この中央値検定は，前提条件が特に必要とされないため汎用性が高いと言えるが，逆に検定力はU検定と比較して非常に低い（Siegel，1956）。したがって，独立する2標本間の中央値の差の検定としては，U検定の適用を第一として考え，散布度の等質性等の前提条件が満たされない場合のみ，その代用として中央値検定を用いる方が適切であろう。

引用文献

岩原信九郎　1979　新しい教育・心理統計 ノンパラメトリック法 第9版 日本文化科学社.

岩原信九郎　1984　新訂版 教育と心理のための推計学 第28版　日本文化科学社.

岩原信九郎　1985　増補版 推計学による新教育統計法 第43版　日本文化科学社.

Kasuya E. 2001 Mann-Whitney U test when variances are unequal. *Animal Behaviour*, 61, 1247-1249.

森敏昭, 1999　Q15　検定の前提条件　繁桝算男・柳井晴夫・森敏昭（編著） Q&A で知る 統計データ解析 DOs and DON'Ts.　サイエンス社, 28－30.

Siegel,S. 1956 *Nonparametric statistics for the behavioral sciences: international student edition*. McGraw-Hill: NY.

Siegel, S. & Castellan, N. J. 1988 *Nonparametric statistics for the behavioral sciences: 2nd edition*. McGraw-Hill: NY.

田中豊, 垂水共之　1999　Windows 版　統計解析ハンドブック　ノンパラメトリック法　共立出版.

参考文献

肥田野直・大川信明・瀬谷正敏　1961　心理教育統計学　培風館

森敏昭・吉田寿夫（編著）　1990　心理学のためのデータ解析テクニカルブック　北大路書房

芝祐順・南風原朝和　1990　行動科学における統計解析法　東京大学出版会

芝祐順・渡部洋・石塚智一（編）　1984　統計用語辞典　新曜社

山内光哉　1987　心理・教育のための統計法　サイエンス社