

Uの分布 (Mann – Whiney U test)

m 個のデータ、 X_1, \dots, X_m 、と、 n 個のデータ、 Y_1, \dots, Y_n があるとき、 D_{ij} を次式で与える (Gibbons, 1971)。

$$D_{ij} = \begin{cases} 1 & X_i > Y_j \\ 0 & X_i < Y_j \end{cases}$$

この D_{ij} の和として U を定義する。

$$U = \sum_{i=1}^m \sum_{j=1}^n D_{ij} \quad (1)$$

すなわち、 U は、 X_i と Y_j を比べたとき、 X_i の方が Y_j より大きい組み合わせの数である。 (1) 式によって与えられる U の値が u である確率を、 $p_{m,n}(u)$ で表わす。

$$p_{m,n}(u) = P(U = u)$$

U の値は、 m 個の文字 X と n 個の文字 Y を混ぜて一直線に並べたときに、これらの mn 個の文字の並びから 2 つの文字（隣り合う文字である必要はない）をとったときに、左側の文字が X で右側の文字が Y である組み合わせの対の総数とみなすことができる。つまり、大きいものから順番に、左から並べられているとみなす。 m 個の X と n 個の Y の並べ方は $\binom{m+n}{m}$ 通りあるが、 X_i と Y_j の分布が同じであるという仮説のもとでは、これらの並べ方の生起確率はいずれも等しいと考えられる。従って、 m 個の X と n 個の Y を並べたときに、 U の値が u になる並べ方の総数を $r_{m,n}(u)$ とおけば、

$$p_{m,n}(u) = \frac{r_{m,n}(u)}{\binom{m+n}{m}}$$

となる。

X と Y の文字の並べ方において、左端が Y のものと X のものとに分けて考えれば、次の関係式の成り立つことがわかる。

$$r_{m,n}(u) = r_{m,n-1}(u) + r_{m-1,n}(u-n) \quad (2)$$

$r_{m,n-1}(u)$ は、左端が Y であるときの並べ方の数を表わす。この場合は、左端の Y を無視して、(n - 1) 個の Y があると考えて数える。

$r_{m-1,n}(u-n)$ は、左端が X であるときの並べ方の数を表わす。(m - 1) 個の X と、n 個の Y を並べてから、左端に X を付け加えると、u の値は n 増えることに注意して、数える。

式 (21) より、 $p_{m,n}(u)$ についての次の関係式が得られる。

$$\begin{aligned} p_{m,n}(u) &= \frac{r_{m,n}(u)}{\binom{m+n}{m}} \\ &= \frac{r_{m,n-1}(u)}{\binom{m+n}{m}} + \frac{r_{m-1,n}(u-n)}{\binom{m+n}{m}} \\ &= \frac{n}{m+n} \cdot \frac{r_{m,n-1}(u)}{\binom{m+n-1}{m}} + \frac{m}{m+n} \cdot \frac{r_{m-1,n}(u-n)}{\binom{m-1+n}{m-1}} \\ &= \frac{np_{m,n-1}(u) + mp_{m-1,n}(u-n)}{m+n} \end{aligned}$$

確率 $p_{m,n}(u)$ は、上式により再帰的に求めることができる。

分布関数は、次式により与えられる。

$$P(u \leq U) = \sum_{u=0}^U p_{m,n}(u) \quad (2)$$

プログラム PCumU.dpr は、U の累積分布確率 (2) 式の値を求めるものである。プログラムを起動すると、図 1 のフォームが表示される。

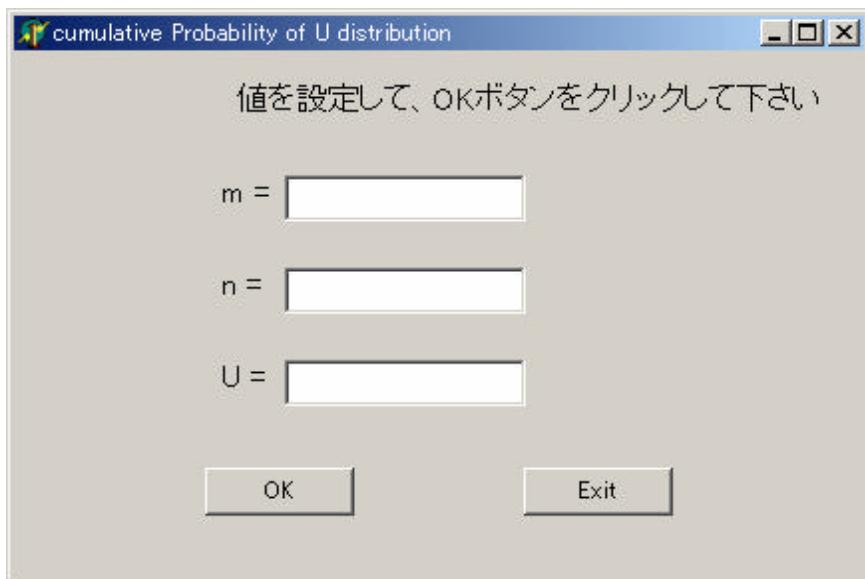


図1 起動時のフォーム

m、n、Uの値を、「m =」、「n =」、「U =」の値として図2のように設定する。

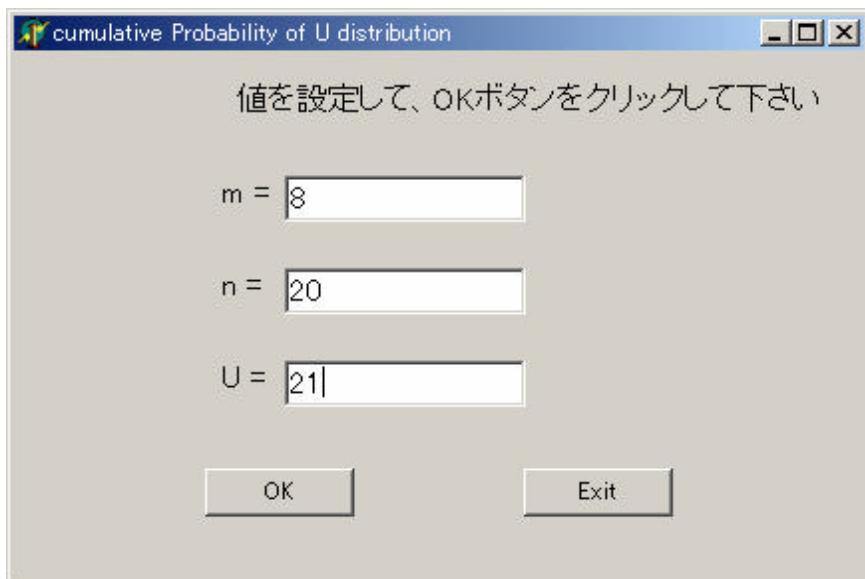


図2 値の設定

値の設定後、「OK」ボタンのクリックで計算が始り、計算結果が図3のように表示される。

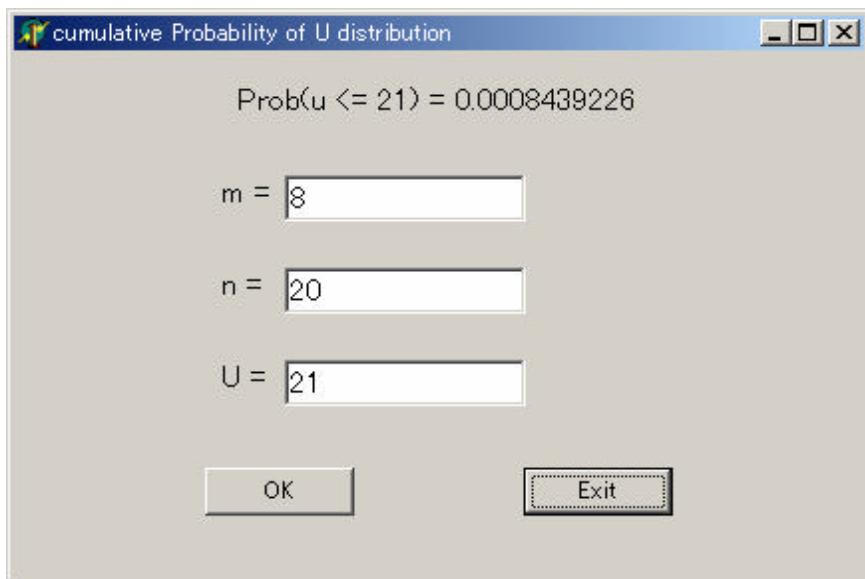


図3 計算結果の表示

図3では、累積確率

$$P(u \leq 21) = \sum_{u=0}^{21} p_{8,20}(u) \approx 0.00084$$

が表示されている。

参考文献

Gibbons,J.D.(1971). *Nonparametric statistical inference*. McGraw-Hill,Inc.

岡本安晴 (1998). Delphi で学ぶデータ分析法 . CQ 出版社 .

Siegel,S (1956). *Nonparametric Statistics: For the Behavioral Sciences*. McGraw-Hill Book Company,Inc.

Siegel,S. & Castellan,N.J.,Jr.(1988). *Nonparametric statistics for the behavioral sciences. 2nd. Ed.* McGraw-Hill,Inc.