

Pearson および Spearman の相関係数に対する p 値を

Fisher の z 変換近似で算出するのに

JMP スクリプト言語を利用する方法について

SAS Institute Japan 株式会社

JMP ジャパン事業部 テクニカルグループ テスター

小野裕亮 (Yusuke.Ono@jmp.com)

注意

Fisher の z 変換近似にはいくつかの方法があります。この資料で説明しているのは、バイアス項なしで、かつ、分散を $1/(n-3)$ としている方法です。また、簡単のため、帰無仮説はゼロ仮説 ($H_0: \rho = 0$) として、かつ、両側検定としています。

はじめに

この資料では、Pearson および Spearman の相関係数に対する p 値を Fisher の z 変換近似で算出するのに、JMP スクリプト言語を利用する方法について紹介いたします。以後、「Pearson および Spearman の相関係数に対して Fisher の z 変換近似で求めた p 値」を、この資料では略して、「z 変換 p 値」と呼びます。

最初に、JMP スクリプト言語全般について、その概略と、簡単な例（正規分布の累積確率を求める例）について説明します。続いて、JMP スクリプト言語を用いた z 変換 p 値の簡単な計算例について解説します。最後に、最後に、より汎用的に z 変換 p 値を求めるプログラムの一例を紹介します。

JMP スクリプト言語について

JMP スクリプト言語の概略

JMP は、マウス操作によって対話的に分析を進める統計ソフトウェアです。そのため、多くのユーザーの方は GUI の機能だけを用いています。しかし、次のような場合に、プロ

プログラミングを行いたい場合があります。

- 似たような処理を繰り返し行う場合
- JMP で備わっていない統計分析や計算を自分自身でプログラミングしたい場合

プログラミングを行う機能として、JMP では、JMP スクリプト言語 (JSL; JMP Scripting Language) と呼ばれているスクリプト言語が提供されています。

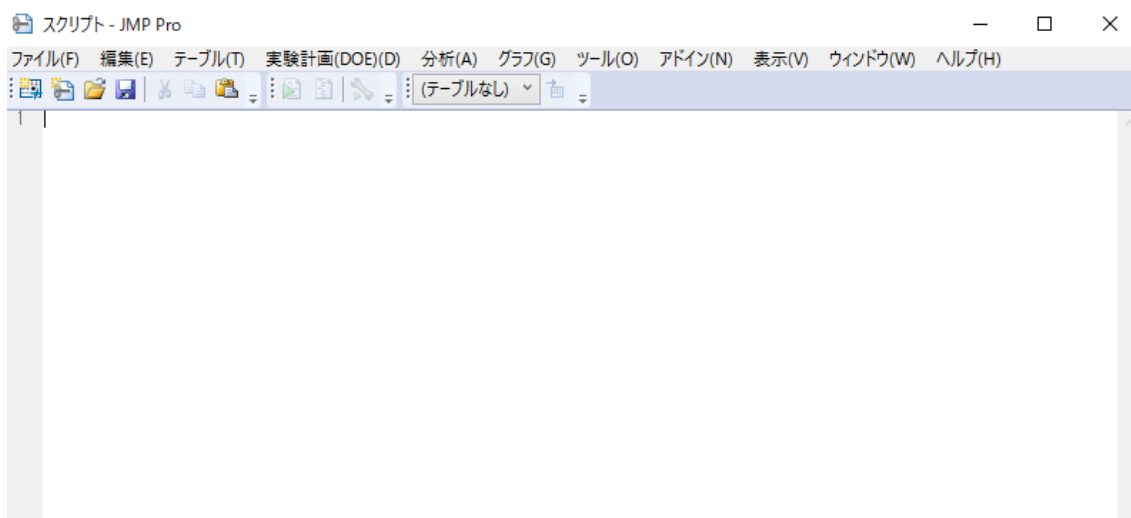
JMP スクリプト言語は、大別すると、「スクリプト」の機能と、「プログラミング言語」の機能に分けることができます。スクリプトとしては、JMP 上でのマウス操作処理が JMP スクリプト言語によって (マウス操作なしに) 実行することができます。プログラミング言語としては、四則演算・確率計算・行列計算・反復処理 (for ループ)・条件処理 (if 文) などによるプログラミングが行えます。

JMP スクリプト言語の簡単な使用例

最初の例として、「標準正規分布に従う確率変数が-2 以下になる確率」を JMP スクリプト言語で求めてみましょう。

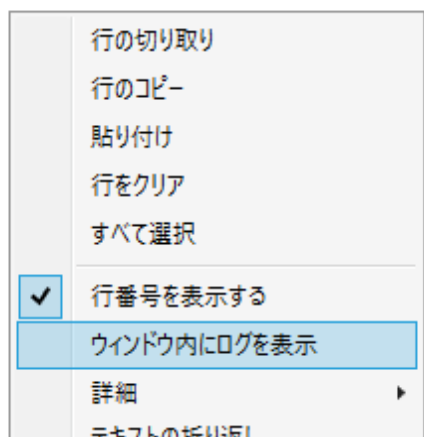
1. スクリプトエディタを呼び出す。

JMP スクリプト言語のコードを書くには、スクリプトエディタを利用するのが便利です。スクリプトエディタを呼び出すには、JMP を起動して、画面上部のプルダウンメニューより [ファイル] → [新規作成] → [スクリプト] を選択します。下図のようなスクリプトエディタが呼び出されます。



2. このスクリプトエディタ内にログウィンドウを表示しておく、エラーを見たり、途中結果を見たりと、何かと便利です。それには、上図のスクリプトエディタ上で右クリッ

クして、呼び出されたメニューより、[ウィンドウ内にログを表示] を選択します（次図）。



- これでプログラミングする準備が整いました。JMPにて、正規分布の累積確率を元円る関数は、Normal Distribution という関数です。この関数の第1引数に-2と指定すると、標準正規分布に従う確率変数が-2以下になる確率が求められます。また、Show関数を指定すると、変数内に保有されている数値がログウィンドウに出力されます。そこで、Normal Distribution関数とShow関数を使って、次のようにプログラミングすると、標準正規分布に従う確率変数が-2以下になる確率がログに出力されます。

```
p = Normal Distribution (-2);  
Show(p);
```

- 上記のコードを実行するには、スクリプトエディタ上で右クリックして、[スクリプトの実行] を選択します。次図のように、ログに結果が出力されます。

```
スクリプト - JMP Pro
ファイル(F) 編集(E) テーブル(T) 実験計画(DOE)(D) 分析(A) グラフ(G) ツール(O) アドイン(N)
表示(V) ウィンドウ(W) ヘルプ(H)
(テーブルなし)
1 p = Normal Distribution(-2);
2 Show(p);
/*:
/*:
/*:
p = Normal Distribution(-2);
Show(p);
/*:
p = 0.0227501319481792;
```

JMP スクリプト言語を用いた z 変換 p 値の計算例

z 変換 p 値の計算式

標本相関係数 r と母相関係数 ρ に対する次式による変換、Fisher の z 変換と呼ばれています。

$$z = \frac{1}{2} \log \left(\frac{1+r}{1-r} \right)$$

$$\zeta = \frac{1}{2} \log \left(\frac{1+\rho}{1-\rho} \right)$$

Fisher の z 変換を施すと、平均が ζ 、分散が $1/(n-3)$ の正規分布に、z が近似的に従います。Fisher の z 変換をしたほうが、標本相関係数 r を何も変換しないで行った正規近似よりも、かなり近似が良くなります。なお、Fisher の z 変換にはいくつかの近似方法が提案されています。上記の方法はバイアス項を含めず、分散を $1/(n-3)$ とした近似になっています。

いま、帰無仮説： $H_0: \rho = 0$ を両側検定するとします。この帰無仮説のもとでは、

$$\sqrt{n-3}z \sim N(0, 1^2)$$

に従います。よって、この両側検定の p 値（近似 p 値）は、

$$p = 2\Phi(-|\sqrt{n-3}z|)$$

となります。ここで、 $\Phi(\cdot)$ は、標準正規分布の下側累積分布関数です。

なお、正確な p 値（ただし、正確になるのは、母集団分布が多変量正規分布のときに限る）は、 $t = -\left| r \sqrt{\frac{n-2}{1-r^2}} \right|$ として、

$$p = 2F_t(t, n-2)$$

から計算されます。ここで、 $F_t(t, n-2)$ は、自由度 $n-2$ の Student の t 分布に従う確率変数が t 以下になる確率（自由度 $n-2$ の Student の t 分布の下側累積分布関数）です。

z 変換 p 値の JMP スクリプト言語による簡単な計算例

前節の計算式に従い、 $H_0: \rho = 0$ の両側検定に対する z 変換 p 値を求める場合、次のようなコードとなります。

```
r = 0.4;  
n = 15;  
z = 0.5 * Log((1+r)/(1-r));  
p_z = 2*Normal Distribution(-Abs(z*Sqrt(n-3)));  
Show(p_z);
```

これは、標本相関係数が 0.4 で、標本サイズが 15 のときの $H_0: \rho = 0$ の両側検定に対する z 変換 p 値を求めています。

3行目の「`z = 0.5 * Log((1+r)/(1-r));`」にて、標本相関係数を z 変換しています。Log 関数は自然対数を求めるものです。

そして、前述の $p = 2\Phi(-|\sqrt{n-3}z|)$ という p 値を、4行目の「`p_z = 2*Normal Distribution(-Abs(z*Sqrt(n-3)));`」にて求めています。Abs 関数は絶対値を、Sqrt 関数は平方根を求めるものです。また、Normal Distribution 関数は、標準正規分布の下側累積分布関数です。

最後の5行目で、結果をログに出力しています。

$H_0: \rho = 0$ の両側検定に対する正確な p 値を求めたい場合には、次のようなコードになりま

す。

```
r = 0.4;  
n = 15;  
t = r * Sqrt((n-2)/(1-r^2));  
p_t = 2 * T Distribution(-Abs(t), n - 2);  
Show(p_t);
```

先ほどと異なって、T Distribution 関数を用いています。この関数は、Student の t 分布の下側累積分布関数です。

この例では、 $p_z=0.1422$, $p_t=0.1396$ となっており、両者の結果はやや大きく違います。

z 変換信頼区間の計算式

前述のように、

$$z = \frac{1}{2} \log \left(\frac{1+r}{1-r} \right)$$

のように Fisher の z 変換を施すと、平均が ζ 、分散が $1/(n-3)$ の正規分布に、z が近似的に従います。ここで $\zeta = \frac{1}{2} \log \left(\frac{1+\rho}{1-\rho} \right)$ です。よって、 $\Pr \left(\zeta - \sqrt{n-3} z_{\frac{\alpha}{2}} < z < \zeta + \sqrt{n-3} z_{\frac{\alpha}{2}} \right) \approx 1 -$

α となります。ここで $z_{\frac{\alpha}{2}}$ は、標準正規分布の上側 $\frac{\alpha}{2}$ 点です。これを ρ について解くと、信頼係数が約 $1 - \alpha$ である信頼区間が求まります。

計算して得られる信頼区間は、次式のようなものです。

$$\left(\frac{\exp(2z_{low}) - 1}{\exp(2z_{low}) + 1}, \frac{\exp(2z_{up}) - 1}{\exp(2z_{up}) + 1} \right)$$

ここで、 $z_{low} = z - z_{\alpha/2}/\sqrt{n-3}$ および $z_{up} = z + z_{\alpha/2}/\sqrt{n-3}$ です。

z 変換信頼区間の JMP スクリプト言語による簡単な計算例

z 変換信頼区間は、次のようなコードで計算されます。

```
r = 0.4;  
n = 15;  
alpha = 0.05;
```

```

z = 0.5 * Log((1+r)/(1-r));
up_z = z + Normal Quantile(1-alpha/2) * Sqrt(1/(n-3));
low_z = z - Normal Quantile(1-alpha/2) * Sqrt(1/(n-3));
Show((Exp(2*low_z)-1)/(Exp(2*low_z)+1));
Show((Exp(2*up_z)-1)/(Exp(2*up_z)+1));

```

5行目と6行目で、Normal Quantile 関数を使っています。この関数は、正規分布の分位点を戻す関数です。

JMP スクリプト言語を用いたより汎用的なプログラム例

前節のコードにより、z 変換 p 値を求めることができます。しかし、いちいち、1 つずつ標本サイズ (n) と標本相関係数 (r) を指定しないといけないため、複数の標本相関係数があると指定が面倒です。

JMP スクリプト言語では、列を指定するための起動ダイアログを Column Dialog という関数で作成できます。また、「多変量の相関」プラットフォームなどの数値的結果を取り出すことができます。

以下のコードでは、指定された列に基づき「多変量の相関」プラットフォームを実行し、そのレポートの「相関の信頼区間」表に、「z 変換 p 値」という列を追加します。また、Spearman の相関係数の表に、「z 変換 p 値」・「z 変換下側」・「z 変換上側」という列を追加します。これらの p 値や信頼区間は、前節までで説明した近似式によって求められています。

なお、このコード中の計算では、

```

z = 0.5 * log((1+r):(1-r));
p_z = 2 * Normal Distribution(-Abs(Sqrt(n - 3):* z));

```

のように、割り算と掛け算にコロン (:) が付いています。このコロンは、行列の要素通しの割り算や掛け算を意味します。変数 r や z は列ベクトルであり、それらに対して要素通しの掛け算や割り算を行っています。

/*

Fisherのz変換による近似p値を
JMPスクリプト言語で求める例

作成者 : Yusuke. Ono@jmp.com (小野裕亮)

このプログラムはサンプルとして提供しているものであり、
従来のテストや検証をまったく行っていません。

```
*/
```

```
// このプログラムで用いる変数名が、データテーブルの変数名と  
// 同じになったときに、名前の衝突を避ける
```

```
Names Default To Here( 1 );
```

```
// 分析対象の列を指定する列ダイアログを呼び出す。
```

```
colinfo = Column Dialog(  
    ylist = ColList( "Y",  
        Min Col( 2 ),  
        Modeling Type( {"Continuous"} )  
    ),  
    H List( Text Box( "有意水準" ), alpha = EditNumber( .05 ) )  
);
```

```
// 「キャンセル」されたときの処理
```

```
If( N Items( colinfo["ylist"] ) <= 1,  
    Throw( "キャンセルされたか、2列以上指定されていません。" )  
);
```

```
alpha = Num( colinfo["alpha"] );
```

```
If( Is Missing( alpha ), alpha = 0.05 );
```

```
// 「多変量の相関」プラットフォームの呼び出し
```

```
plat = Multivariate(  
    Y( colinfo["ylist"] ),  
    CI of Correlation( 1 ),  
    Pairwise Correlations( 1 ),  
    Correlation Probability( 1 ),  
    Spearman's  $\rho$ ( 1 ),
```



```

        Set  $\alpha$  Level(alpha)
    );

// レポートから結果を取り出す */
rep = Report(plat);
/* ペアで両方とも非欠測値となっている個数 */
n = rep[Outline Box("ペアごとの相関")][Number Col Box("度数")]<<Get;
/* 相関係数推定値 */
r = rep[Outline Box("相関の信頼区間")][Number Col Box("相関")<<Get;

/* z変換によるp値の計算 */
z = 0.5 * log((1+r)/(1-r));
p_z = 2 * Normal Distribution(-Abs(Sqrt(n - 3) * z));

/* 「相関の信頼係数」の表に、上記のp値を追加 */
rep[Outline Box("相関の信頼区間")][Table Box(1)]<<Append(ncb_zp = Number Col Box("z変換p値", p_z));
/* 表示形式の指定 */
ncb_zp << Set Format(9, 97);
ncb_zp << Set Conditional Format("p値");

/* 以下、Spearman */
/* Spearman 相関係数推定値 */
s = rep[Outline Box("ノンパラメトリック: Spearmanの順位相関係数( $\rho$ )")][Number Col Box("Spearmanの順位相関係数( $\rho$ )")<<Get;
/* z変換によるp値の計算 */
z_s = 0.5 * log((1+s)/(1-s));
p_z_s = 2 * Normal Distribution(-Abs(Sqrt(n - 3) * z_s));

/* 「相関の信頼係数」の表に、上記のp値を追加 */
rep[Outline Box("ノンパラメトリック: Spearmanの順位相関係数( $\rho$ )")][Table Box(1)]<<Append(ncb_pzs = Number Col Box("z変換p値", p_z_s));
/* 表示形式の指定 */
ncb_pzs << Set Format(9, 97);

```

```
ncb_pzs << Set Conditional Format("p値");
```

```
z_low = z_s - Normal Quantile(1-alpha/2) :/ Sqrt(n- 3);
```

```
z_up = z_s + Normal Quantile(1-alpha/2) :/ Sqrt(n- 3);
```

```
low = (exp(2 * z_low) - 1) :/ (exp(2 * z_low) + 1);
```

```
up = (exp(2 * z_up) - 1) :/ (exp(2 * z_up) + 1);
```

```
rep[Outline Box("ノンパラメトリック: Spearmanの順位相関係数( $\rho$ )")] [Table
```

```
Box(1)] << Append(sI = Number Col Box("z変換下側", low));
```

```
rep[Outline Box("ノンパラメトリック: Spearmanの順位相関係数( $\rho$ )")] [Table
```

```
Box(1)] << Append(su = Number Col Box("z変換上側", up));
```

```
sI << Set Format(7, 4);
```

```
su << Set Format(7, 4);
```

JMP スクリプト言語の関数やメッセージについては、[ヘルプ] → [スクリプトの索引] を選択すると、簡易的なヘルプとコード例が表示されます。上記の関数やメッセージについてご興味がある場合には、こちらの「スクリプトの索引」でご確認ください。また、スクリプトについてより細かく学んでいきたい場合には、JMP のオンラインヘルプの「スクリプトガイド」をご覧ください。

注意

Fisher の z 変換近似にはいくつかの方法があります。この資料で説明しているのは、バイアス項なしで、かつ、分散を $1/(n-3)$ としている方法です。

また、簡単のため、帰無仮説はゼロ仮説 ($H_0: \rho = 0$) として、かつ、両側検定としています。

-- 以上です。

変更履歴

2020 年 3 月 16 日 初版