

[教材研究]

JMPによる心理統計学入門 An introduction to psychological statistics with JMP

長 潔容江[†]

Kiyoe Cho[†]

[†]久留米大学 心理学研究科

[†] Graduate School of Psychology, Kurume University.

4. はじめに

JMP^{註1}は、世界的に有名な SAS 社によって開発された統計解析ソフトウェアである。新村¹⁾によると、JMP は、SAS 社の創業者の一人である John Sall 博士の設計によって開発された。JMP (ジャンプ) という名前は、「視覚的なデータ分析の世界へジャンプしましょう」というコンセプトからきている。

さまざまな研究領域において、収集したデータは統計的手法によって分析される。心理学の領域においても、データは統計的に処理されることが多い。統計を行う際に用いられる統計用のソフトウェアは実に多種多様であるが、基本的な統計を行う場合には、本稿で分析の際に使用した統計解析用ソフトウェア JMP を使用することをお勧めしたい。その理由として、グラフィック、統計、データのすべてがリンクされていることが挙げられる。この JMP では、統計検定量を算出できるだけでなく、収集したデータをグラフで表現し視覚的に見ることが可能である。このグラフ機能が充実している点が JMP の最大の特徴であると言える。データ収集を終えたら、すぐに統計にかけて結果を出したいところではあるが、その前に集めたデータがどのような分布であるかを目で見て確認することは、非常に重要な作業であると考えられる。

本稿は、心理学を専攻している大学生を対象としている。そのなかでも、心理統計学を学び始めた初学者を対象としている。また、本稿で分析の手法を解説する際に使用しているデータは、筆者の絵画の評価に関する研究のデータの一部である。これは、A~L の 12 種類の絵画作品に関し、絵画の評価を測定する形容詞対 6 項目 (美しい-醜い, 楽しい-寂しい, 好きな-嫌いな, 面白い-つまらない, 良い-悪い, 快い-不快な) について SD 法 (7 段階評定) で回答してもらったデータである。このデータを用いて、入力したデータを JMP に読み込む方法、ヒストグラムの作成、箱ひげ図の見方、散布図の作成と相関分析、そして t 検定までの基本的な心理統計学の手法を解説する。

なお、本稿で使用した JMP のバージョンは、JMP 10 である。JMP 10 を使用している方は、本稿を見ながら問題なく操作を進めることができると思われる。しかしながら、それ以前のバージョンの JMP を使用している方は、表示形式や用いられている言葉などが若干異なる場合があるが、基本的な操作は同じであるため参考にしていただきたい。

註 1

JMP は SAS Institute Inc. の登録商標である。

5. Excel データを JMP で読み込む

2.1 分析したいデータを Excel に入力する

JMP を用いて分析を行う前に、分析したいデータを Excel に入力する。Excel の Sheet の上部には空白を設けず、データの入力の例のように、データは左上から Sub. (調査対象者) ごとに 1 行ずつ入力する。データ入力が終わったら、名前を付けて保存し、Excel を閉じる。

このデータは、A~L の 12 種類の絵画作品の評価 (美しい-醜い, 楽しい-寂しい, 好きな-嫌いな, 面白い-つまらない, 良い-悪い, 快い-不快な) の平均である。

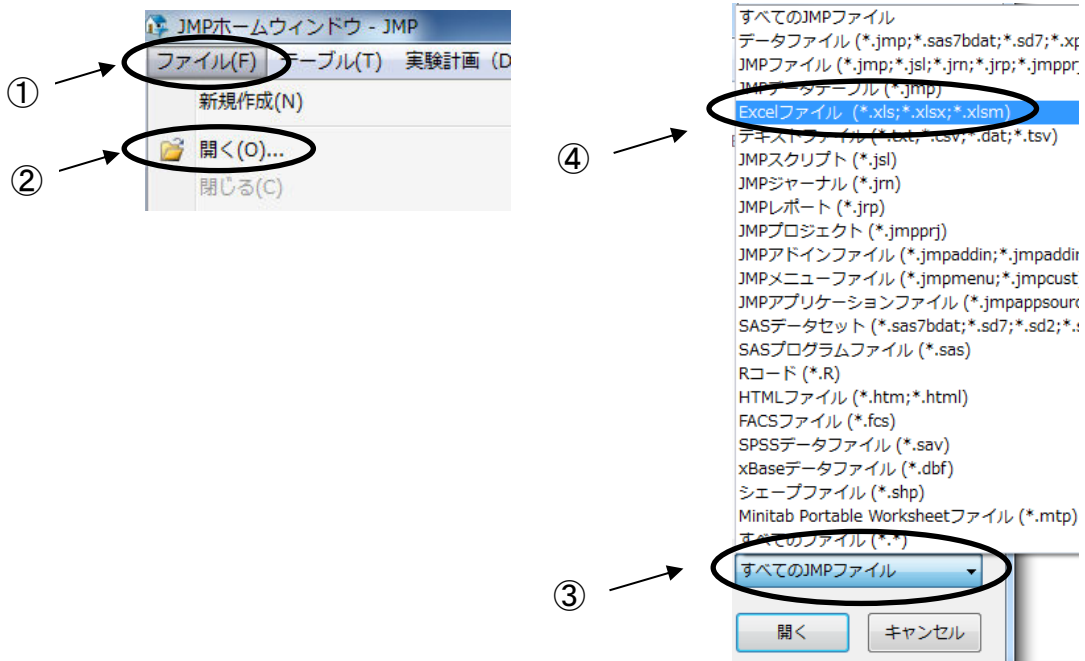
データ入力の例

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Sub.	A	B	C	D	E	F	G	H	I	J	K	L
2	1	2.33	2.33	2.83	2.17	2.50	2.50	3.50	2.67	2.50	2.00	2.33	1.83
3	2	4.00	3.50	3.67	3.33	3.83	4.17	4.50	3.50	3.50	4.83	3.50	4.83
4	3	3.83	3.17	5.17	3.33	4.33	3.83	3.67	3.83	4.00	4.50	3.50	5.00
5	4	4.00	3.83	3.67	3.67	3.33	3.50	4.00	2.83	3.83	4.00	3.83	5.00
6	5	2.67	2.00	2.00	1.83	1.00	1.00	5.17	1.00	1.00	7.00	3.17	7.00
7	6	3.00	4.00	4.17	3.67	3.00	4.00	4.17	3.33	4.17	4.33	4.33	4.17
8
9
10
65	64	4.50	3.33	4.00	3.33	4.00	3.83	3.50	3.50	4.17	5.00	4.17	4.33
66	65	4.17	1.00	2.33	1.33	3.00	1.00	4.33	1.00	1.33	4.33	4.00	4.17

2.2 JMP でデータを読み込む

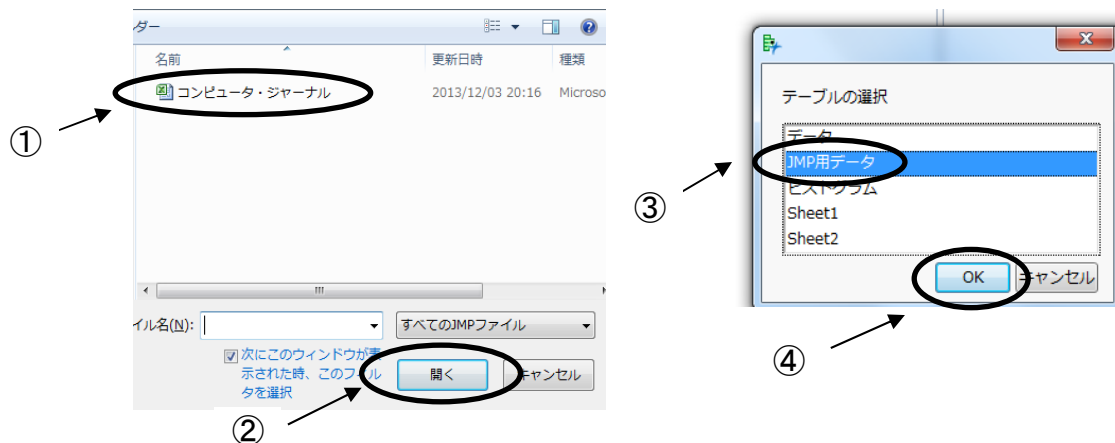
JMP を起動する。その後「ファイル」をクリックし、「開く」をクリックする。

「すべての JMP ファイル」をクリックし、「Excel ファイル」を指定する。



JMP で読み込みたい Excel ファイルをクリックし、「開く」をクリックする。

その後テーブルの選択画面に移るので、分析したいデータを入力した Sheet, この場合「JMP用データ」を選択し、「OK」をクリックする。



6. ヒストグラムでデータの分布を確認する

3.1 ヒストグラムとは

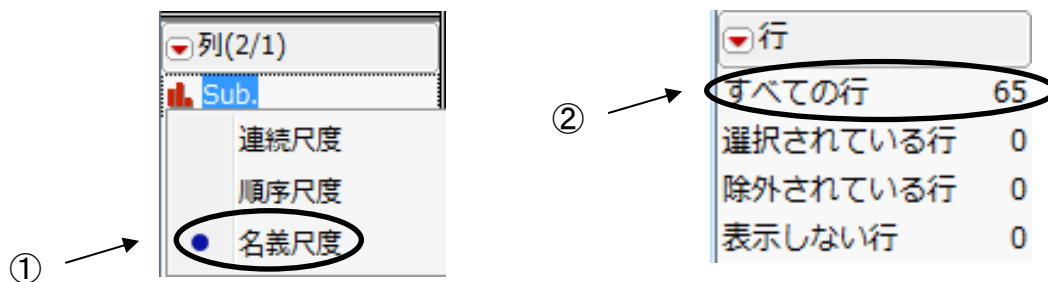
ヒストグラムとは、度数を縦軸、データの区間を横軸に表現した棒グラフである。データの分布を見る場合に、このヒストグラムがよく用いられる。棒と棒の間が隙間なくくっついている理由は、数量データを連続尺度^{註2}とみなしているためである^[1]。

ヒストグラムを作成することによって、データの極端値や分布形のチェックができる。

3.2 ヒストグラムを作成する

JMP は、データが数値の場合、すべて連続尺度とみなしてしまうので、データに応じて尺度を変更する必要がある。この場合、「列」の「Sub.」をクリックし、「名義尺度」にチェックを入れる。

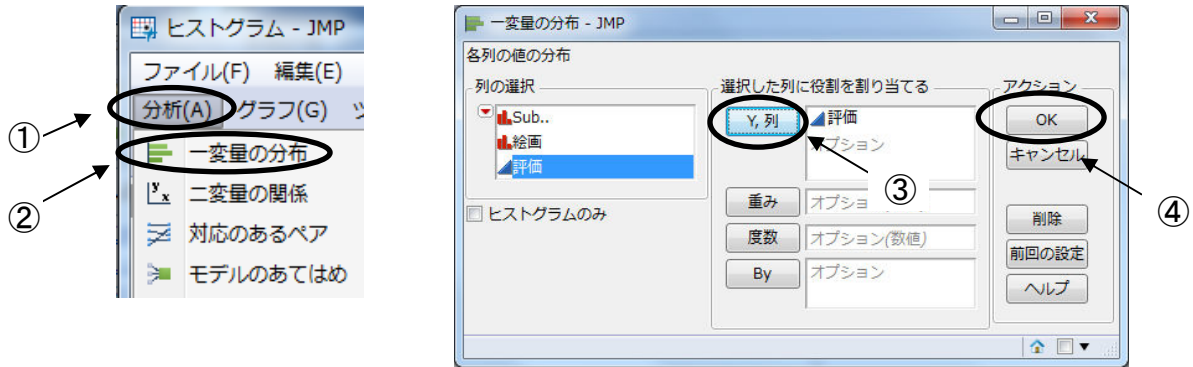
また、「行」の「すべての行」にはデータの総数が表示される。この場合、データ数は 65 個であることが確認できる。



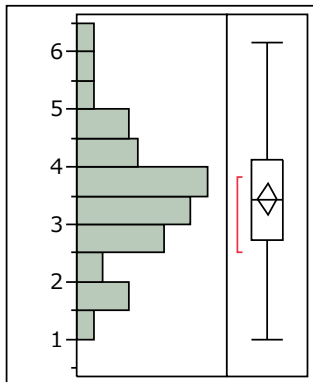
註2

JMP では間隔尺度、比率尺度を合わせて連続尺度と表記している。

「分析」をクリックし、「一変量の分布」をクリックする。
 「列の選択」から、分析したい変数を選択する。ここでは、「評価」について分析するので、「評価」を選択し、「Y、列」をクリック、「OK」をクリックする。

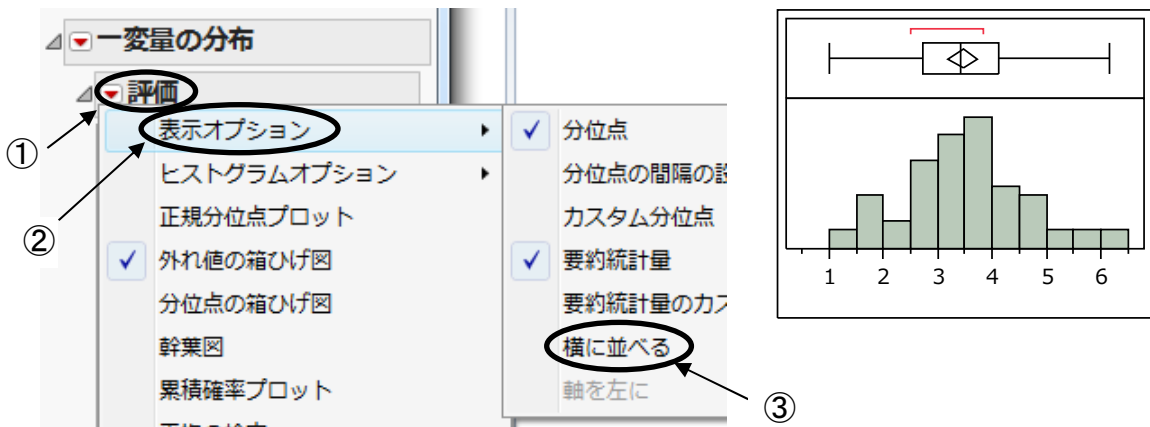


すると、縦向き（縦向き）のヒストグラム、分位点、要約統計量が表示される。
 ヒストグラムを横向き（横向き）に表示するには、「評価」の左側にある▼をクリックし、「表示のオプション」→「横に並べる」をクリックする。
 すると、横向き（横向き）のヒストグラムが表示される。



分位点		
100.0%	最大値	6.17
99.5%		6.17
97.5%		6.0595
90.0%		4.898
75.0%	四分位点	4
50.0%	中央値	3.33
25.0%	四分位点	2.67
10.0%		1.83
2.5%		1.2145
0.5%		1
0.0%	最小値	1

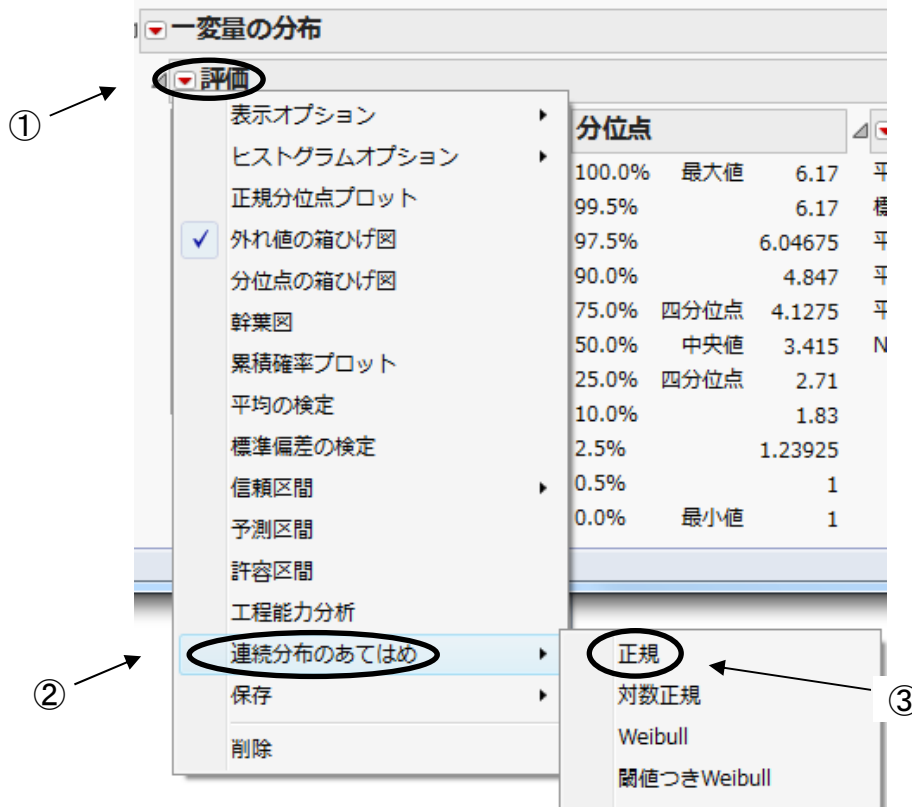
要約統計量	
平均	3.4123077
標準偏差	1.1168423
平均の標準誤差	0.1385272
平均の上側95%	3.6890477
平均の下側95%	3.1355677
N	65



3.3 データは正規分布しているか？（適合度検定）

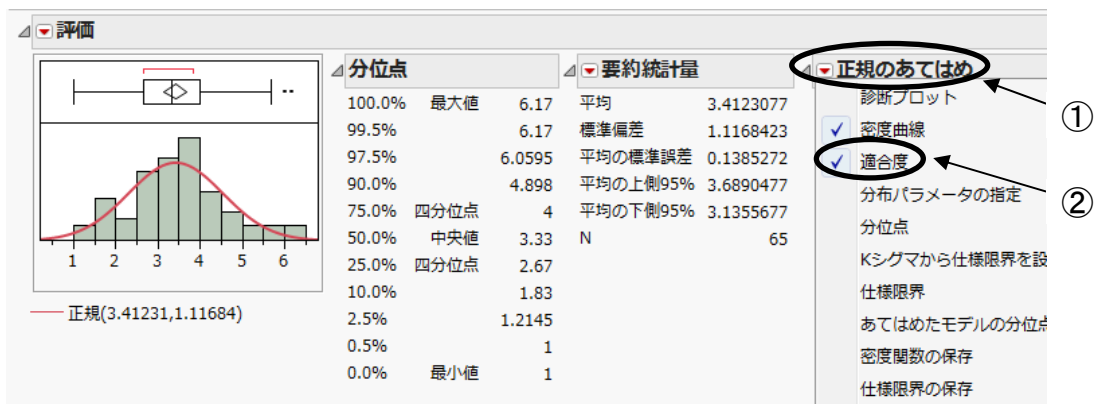
次に、データが正規分布しているかについて、適合度検定を行い確認する。

「評価」の左側にある▼をクリックし、「連続分布のあてはめ」→「正規」をクリックする。



すると、「正規のあてはめ」が表示される。

次に、「正規のあてはめ」の左側にある▼をクリックし、「適合度」をクリックする。

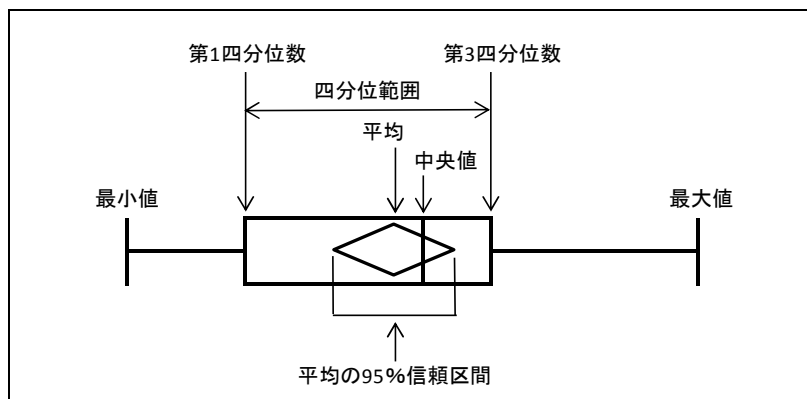


「適合度検定」が表示される。 p 値が 0.05 以上の場合、帰無仮説は棄却できないので有意差なし。つまり、正規分布しているということになる。反対に p 値が 0.05 未満の場合、帰無仮説は棄却され、対立仮説が採用されるので有意差あり。つまり、正規分布していないということになる。このデータは、 $p=0.6087$ であり、正規分布していることが分かる。

適合度検定	
Shapiro-WilkのW検定	
W	p値(Prob<W)
0.984813	0.6087
注: Ho = 正規分布からのデータ。 p値が小さい場合はHoを棄却。	

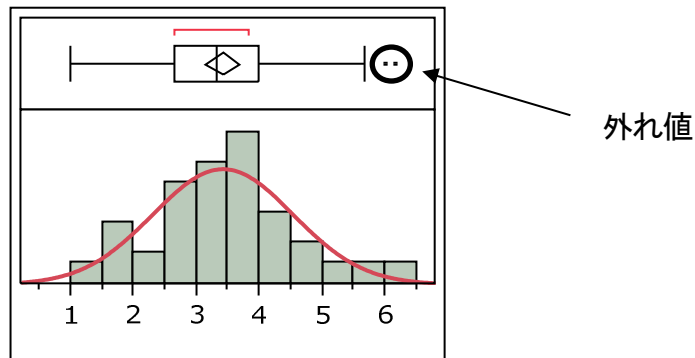
3.4 箱ひげ図

箱ひげ図とは、ばらつきのあるデータを分かりやすく表現するためのグラフである。箱とひげ（線）を用いて、最小値、第1四分位数、平均、中央値（第2四分位数）、第3四分位数、最大値などを表している。



箱ひげ図の見方

こちらが絵画Iの評価の箱ひげ図である。箱ひげ図より、平均が中央値より左側にあり、分布がやや左側に寄っているが、分布の歪みが少ないことが見て分かる。なお、箱ひげ図の最大値の右側にあるドットは、この分布から見た外れ値を示している。



7. 散布図と相関分析

4.1 散布図とは

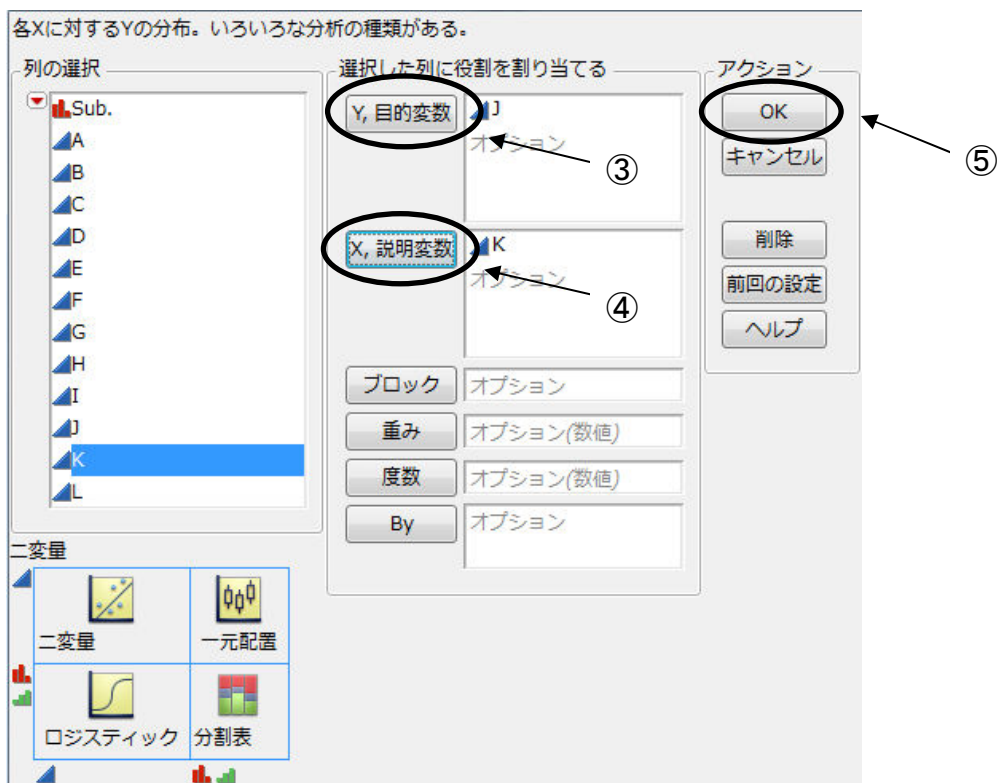
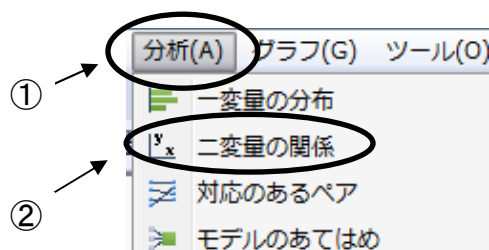
散布図とは、対応する2つの変数の数値の1つを横軸に、もう1つを縦軸にそれぞれ対応させて、平面上に点（ドット）を打って表した図である。散布図を作成することによって、2つの変数の間に関連があるかどうかを、視覚的に確認することができる。

また、相関係数（ r ）は、2つの変数の関係の強さと方向を示す指標であり-1~1までの値をとる。

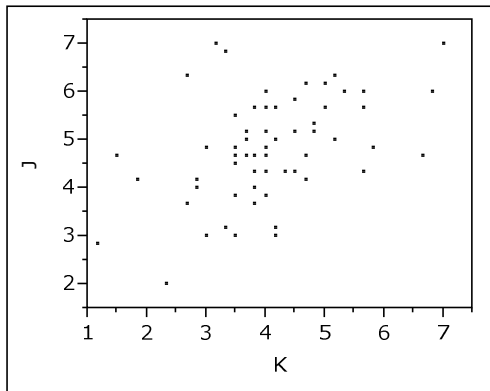
4.2 散布図の描き方

「分析」をクリックし、「二変数の関係」をクリックする。JMPでは、二変量の尺度の組み合わせに応じて、適切な分析方法が選択される。本データはともに連続尺度なので、相関分析が行われる。

ここでは、「J」を「Y, 目的変数」、「K」を「X, 説明変数」に割り当てた場合について説明する。変数を割り当てたら、「OK」をクリックする。



このように散布図が表示される。

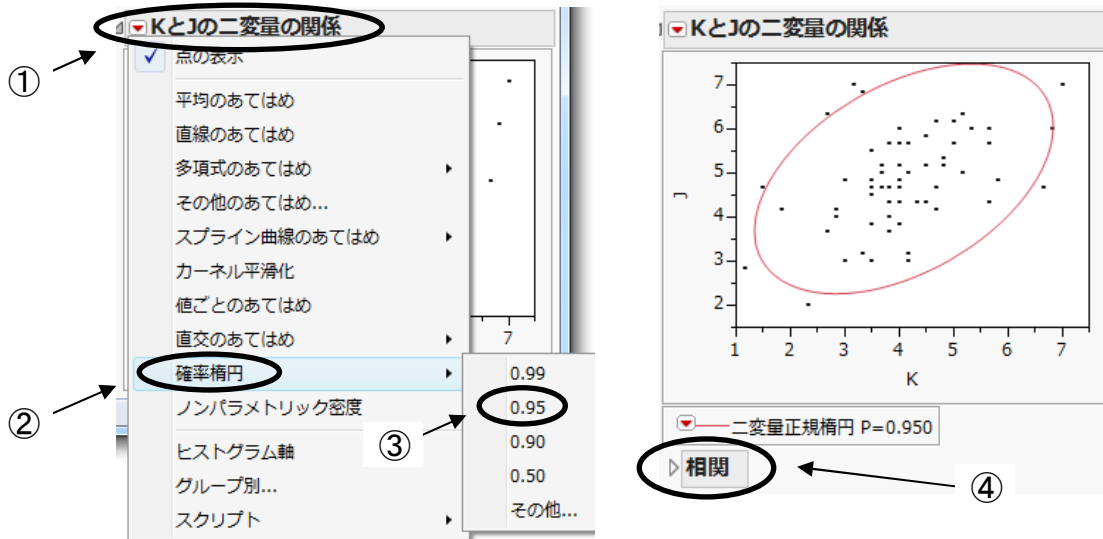


4.3 相関分析を行う

散布図でデータの分布を確認したら、次に相関分析を行う。

相関係数を表示するには、「K と J の二変量の関係」の左側にある▼ボタンをクリックし、「確率楕円」→「0.95」をクリックする。

すると、楕円が表示される。その後、「相関」の左側にある▷ をクリックする。



相関係数が表示される。相関係数 $r = 0.45485$ ($p < .01$) となり、絵画 J と絵画 K の間には、有意な中程度の正の相関がみられることが分かる。

相関					
変数	平均	標準偏差	相関	p値	数
K	4.084923	1.123715	<u>0.45485</u>	<u>0.0001*</u>	65
J	4.874923	1.0664			

8. t 検定

5.1 t 検定とは

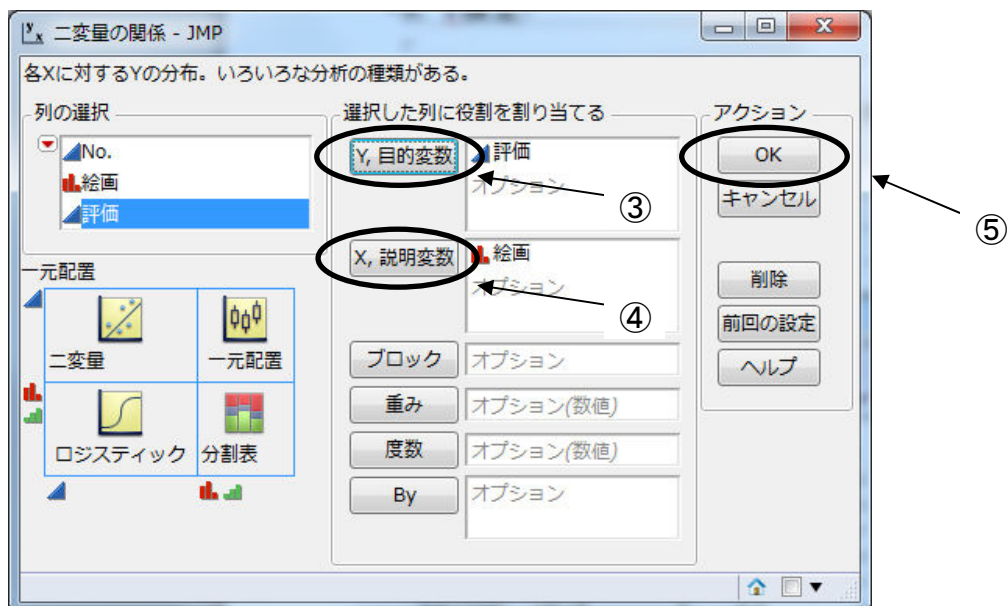
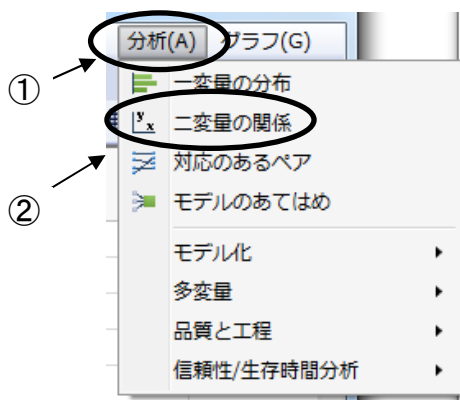
t 検定とは、2つのデータの平均の有意差を分析するものである。なお、分散分析は2つ以上の平均の有意差を検定することができる。平均が2つの場合、 $t^2 = F$ の関係がある。

5.2 t 検定を行う

「分析」をクリックし、「二変量の関係」をクリックする。

ここでは、絵画によって評価が異なるかを見るため^{註3}、「列の選択」の「評価」→「Y, 目的変数」に、「絵画」→「X, 説明変数」に割り当て、「OK」をクリックする。

すると、説明変数が名義尺度、目的変数が連続尺度なので、一元配置の分析が行われる。

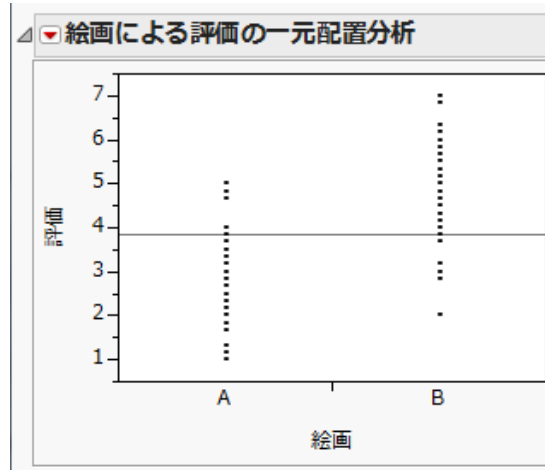


註3

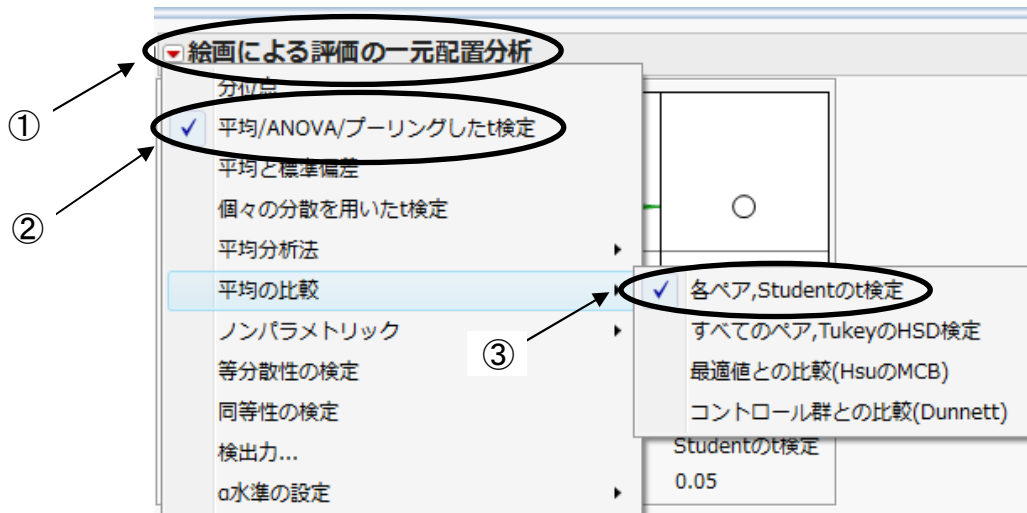
本データの場合、対応のある t 検定を行うべきであるが、本稿では、 t 検定の基本的な分析方法を説明するために、対応のない t 検定を行っている。

この図はドットプロットといい、横軸を独立変数^{註4}、縦軸を従属変数として、データをプロットしたものである。

本データの場合、絵画Aと絵画Bのばらつきはほぼ同じであり、さらに、絵画Bのほうが絵画Aよりも評価の得点が高いことが分かる。



「絵画による評価の一元配置分析」の▼をクリックし、「平均/ANOVA/プーリングした t 検定」をクリックし、「平均の比較」→「各ペア, Studentの t 検定」をクリックする。



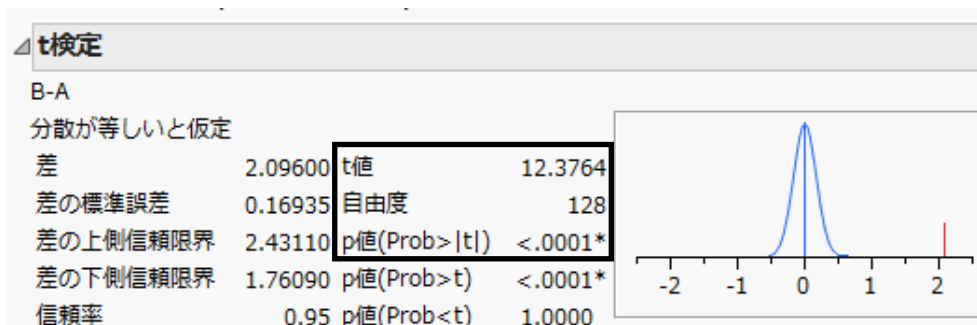
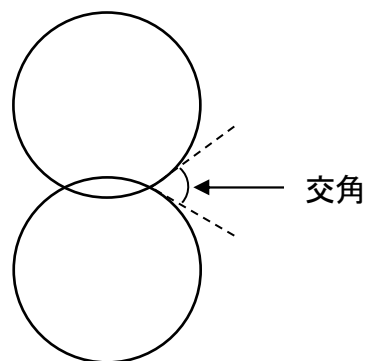
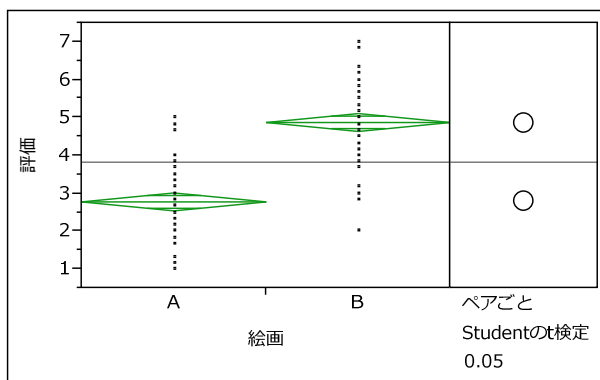
註4

t 検定や分散分析の場合、説明変数は独立変数、目的変数は従属変数と呼ばれる。

グラフにひし形が表示される。左のひし形の中央の直線が絵画 A の平均、右のひし形の中央の直線が絵画 B の平均を示している。ひし形の縦の幅は各グループの母平均の 95%信頼区間を表している。、横幅はデータ数に応じた比率で表示される。

また、右側に比較円がプロットされる。平均の差が有意な場合、比較円は重ならない、あるいは重なり程度の浅く、円の交角は 90° 以下となる。見た目では判断が付きにくい場合には、どちらか一方の比較円をクリックする。すると、有意差がある場合は、もう片方の円の輪郭線が太いグレーで、有意差がない場合は、細い赤で表示される。

t 検定の結果、絵画 A と絵画 B の平均に有意差が見られた ($t(128) = 12.3764, p < .01$)。したがって、絵画 B は絵画 A よりも高く評価されたことが分かる。



9. おわりに

本稿では、上記のような方法で、絵画の評価のデータを用いて、データ入力から t 検定までの手法を解説した。データを分析にかける前には、ヒストグラムを作成してデータの分布を確認する。相関分析を行う前には散布図を描き、さらに楕円を加えてデータの分布を確認する。 t 検定を行う際には、2 つの変数の平均はどれくらい離れているかを一元配置分析のグラフによって確認する。

私たちは、データを収集したら真っ先に統計にかけて結果を出してしまいがちである。しかし、結論を出す前にデータの分布をグラフで視覚的に見るのが重要である。

卒業論文作成などでデータを JMP で分析する際に、本稿が一助となれば幸いである。

参考文献

- [2] 新村秀一, “JMP による統計レポート作成法” 丸善株式会社, 2007.