

統計学(基礎)

第11回 相関

量的データの関連を調べる

相関

統計解析で何ができるか

- ・ データの整理・要約をする
 - 基本統計量、表・グラフ作成
- ・ 比較する・違い(差)を知る
 - 検定
- ・ 関係や傾向を知る ←いまここ
 - 相関・回帰
- ・ グループ分けをする
 - 多変量解析

散布図

- 2つの量的変数のグラフ
 - 同一のケースの2つの変数

	No	DBP	SBP
1	1	78	134
2	2	86	161
3	3	91	146
4	4	75	119
5	5	86	137
6	6	64	135
7	7	78	124
8	8	73	121
9	9	72	136
10	10	80	138

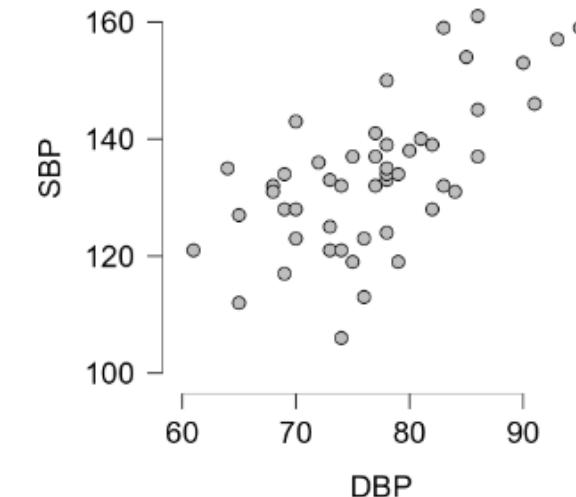
記述統計 ▾

記述統計 ▾

	DBP	SBP
妥当	50	50
欠損値	0	0
平均値	76.90	133.8
標準偏差	7.541	12.47
最小値	61.00	106.0
最大値	95.00	161.0

散布図

DBP - SBP

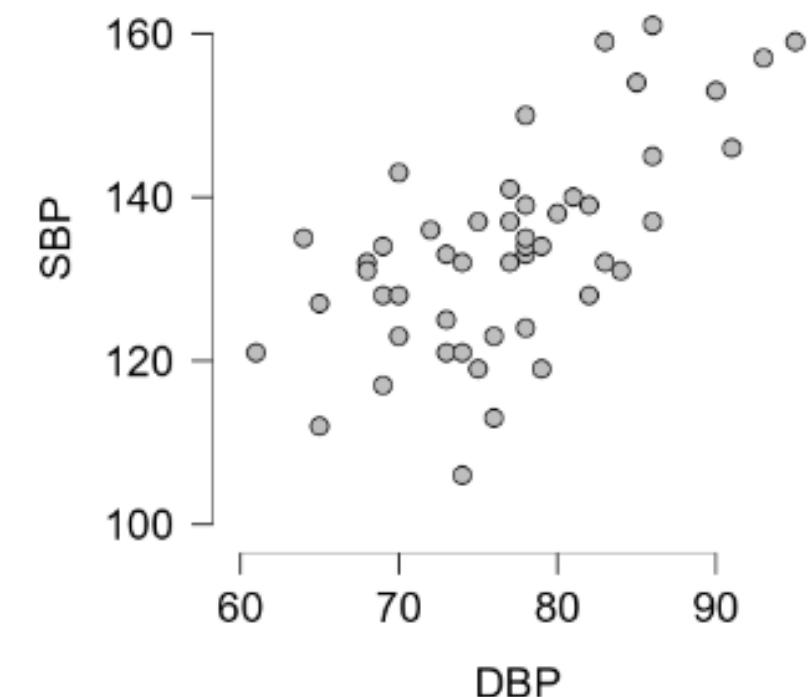


散布図

- 何か関連性が見えないか？

散布図 ▼

DBP - SBP



相関

- 2変数間の関係
 - 量的な変数の関係
 - 順序尺度もできないわけじゃない
 - 直線的な関係
 - 一方が大きくなったときに、もう一方の大小がどうなるか
 - 片方が増えるともう片方も増える
 - 片方が増えるともう片方は減る
 - 片方の増減ともう片方の増減は関係ない

相関係数

- ・ ピアソンの積率相関係数(r)
 - 量的データ同士の関係をみている

- ・ 相関係数の式

$$r = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\{\sum_i (x_i - \bar{x})^2\}\{\sum_i (y_i - \bar{y})^2\}}} = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{(n - 1)s_x s_y}$$

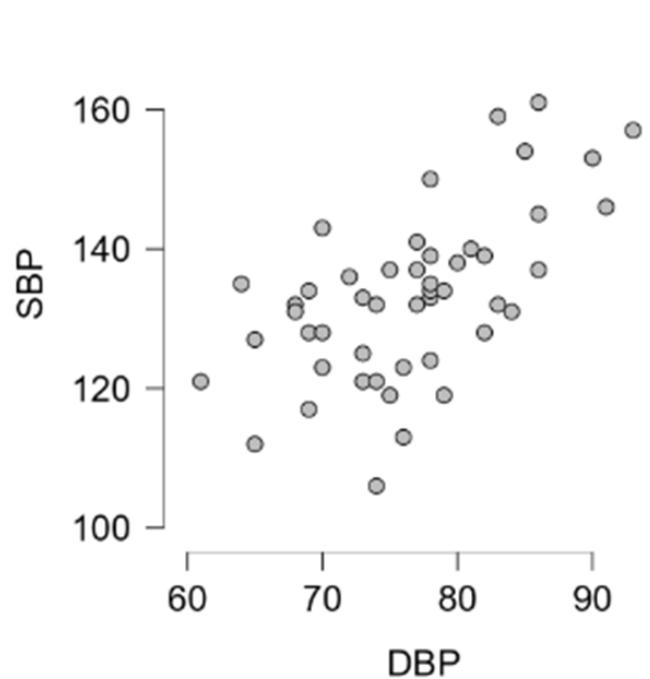
s_x と s_y はそれぞれの(標本)標準偏差

相関

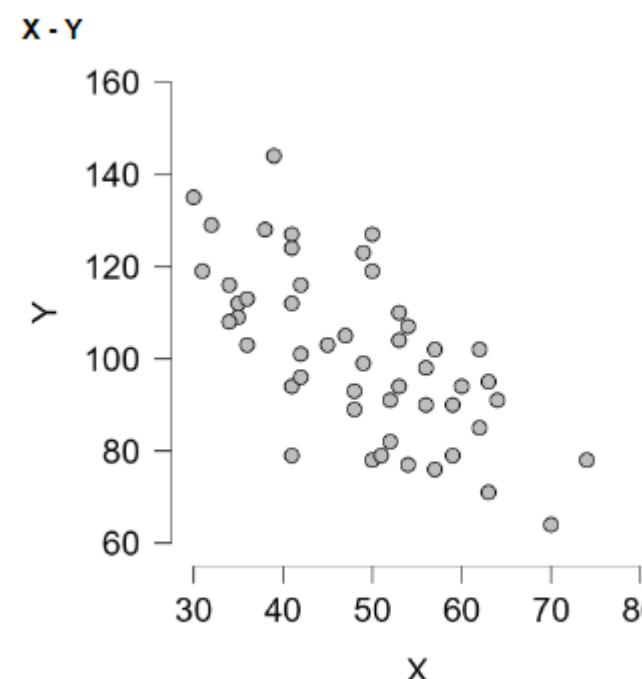
- 正の相関
一方が大きいともう一方も大きい
- 負の相関
一方が大きいともう一方は小さい
- 無相関
一方の大小ともう一方の大小は関係ない

相関のイメージ

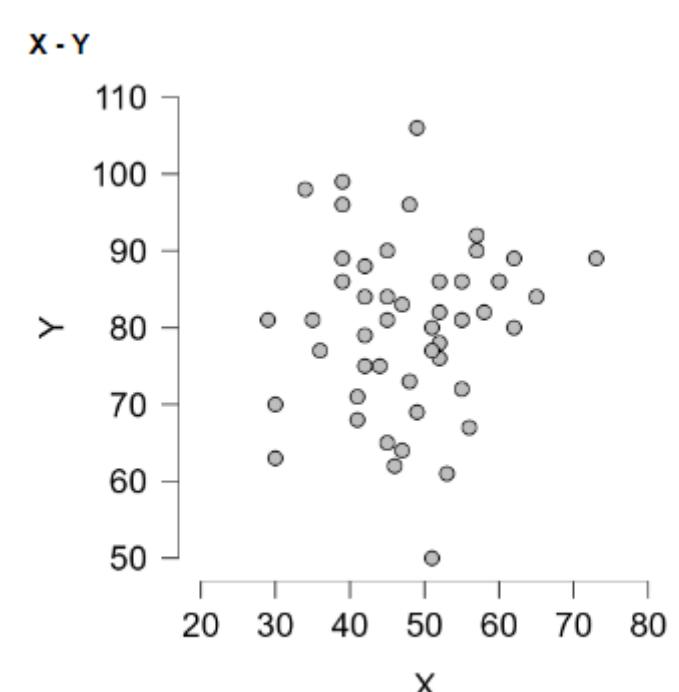
正の相関



負の相関



無相関



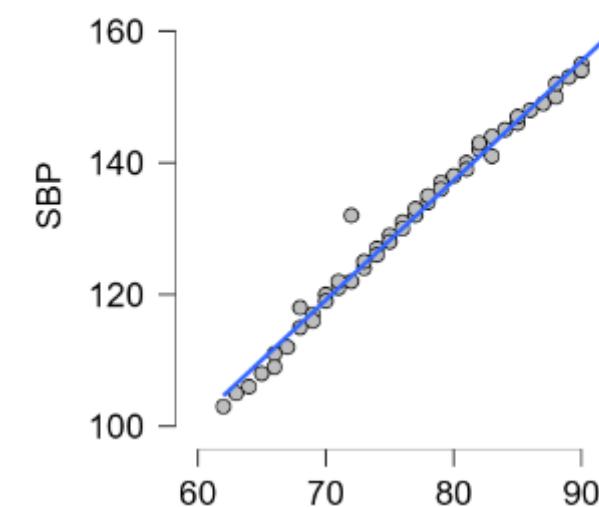
相関係数の解釈 (結構アバウト)

- 0.2より大きい 正の相関
- -0.2より小さい 負の相関
- -0.2から0.2の間無相関
- $|0.2| \sim |0.4|$ 弱い相関
- $|0.4| \sim |0.6|$ 中程度の相関
- $|0.6| \sim |1.0|$ 強い相関

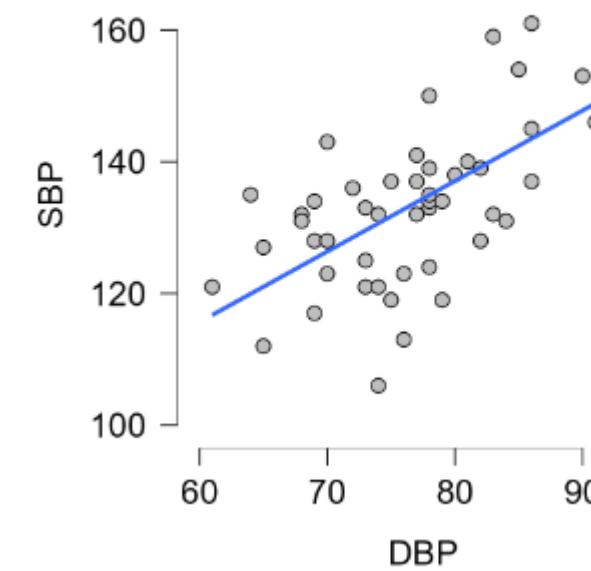


相関のイメージ

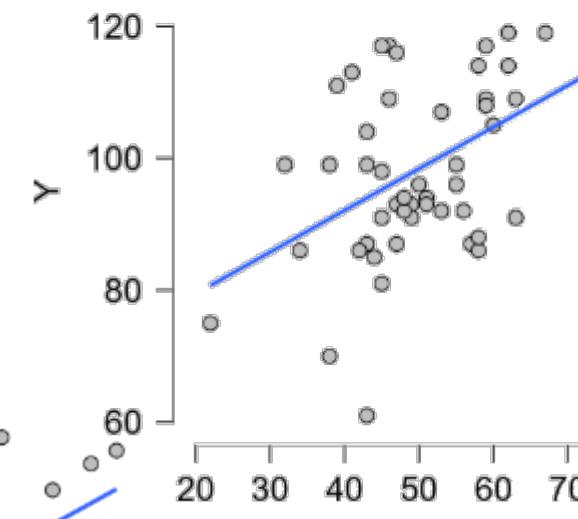
DBP - SBP ▼ $r=0.93$



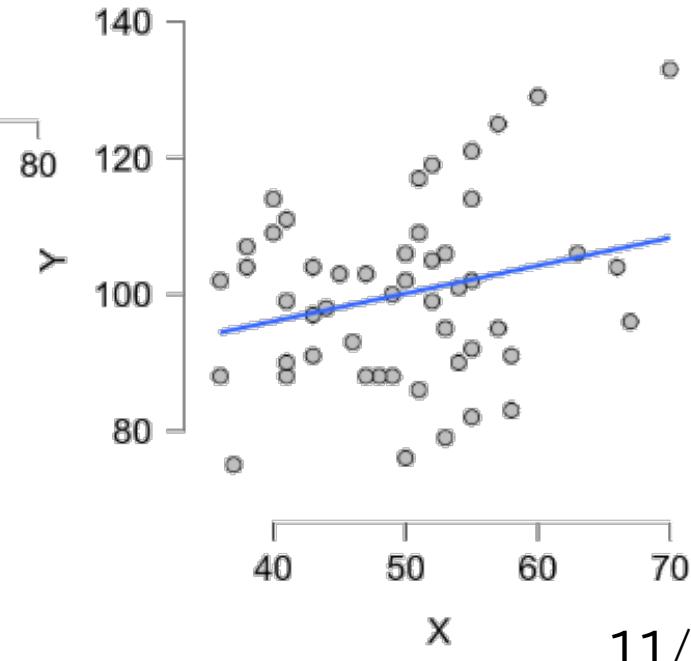
DBP - SBP $r=0.65$



x - Y $r=0.45$



x - Y $r=0.25$

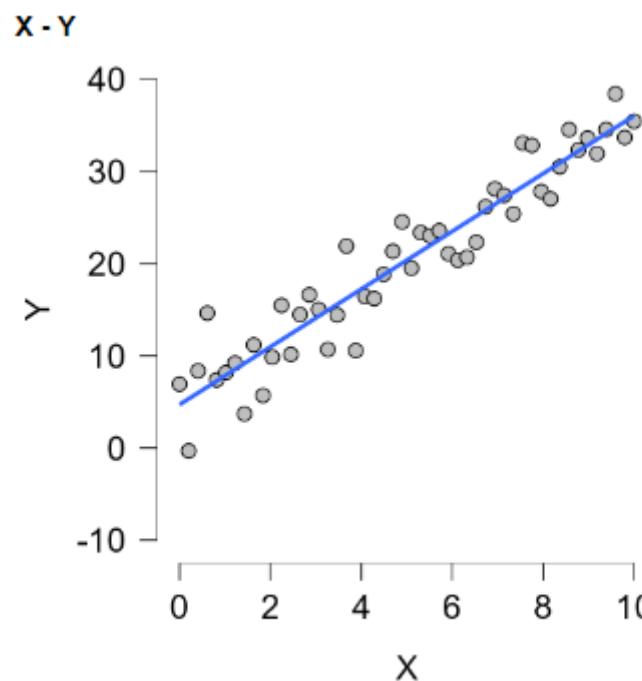


注意点

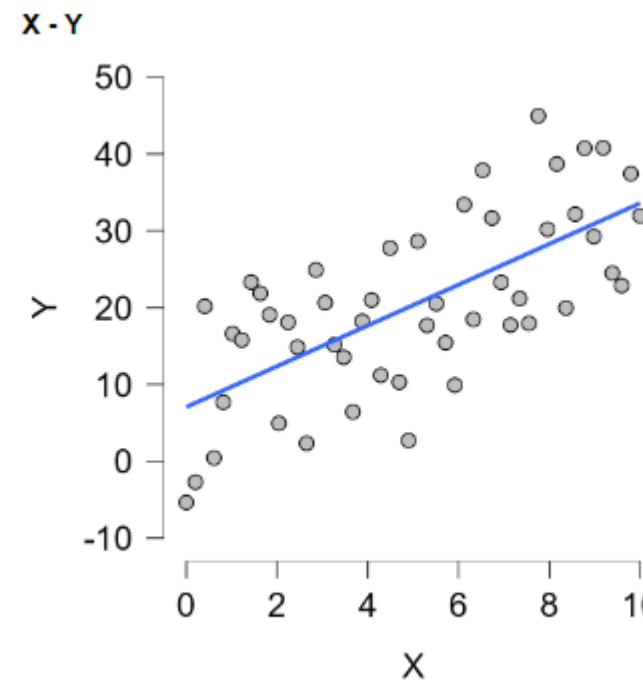
- ・ 散布図を書くときに、2変数で一方を従属、一方を独立変数にした回帰直線を引くことが多い
 - 散らばりの具合を見ている
 - 直接相関係数とは関係ない
- ・ 相関係数の強さと線の傾きの角度は無関係

傾きと相関は無関係

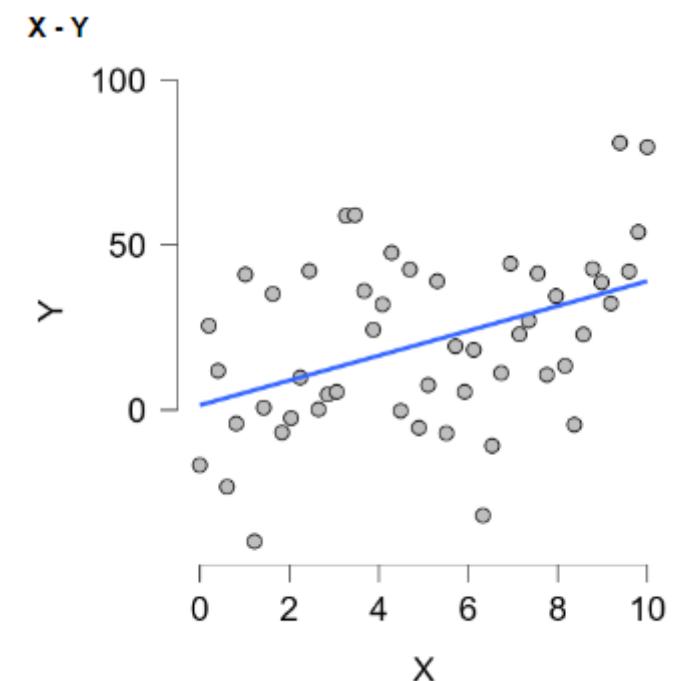
$r=0.95$



$r=0.67$



$r=0.42$

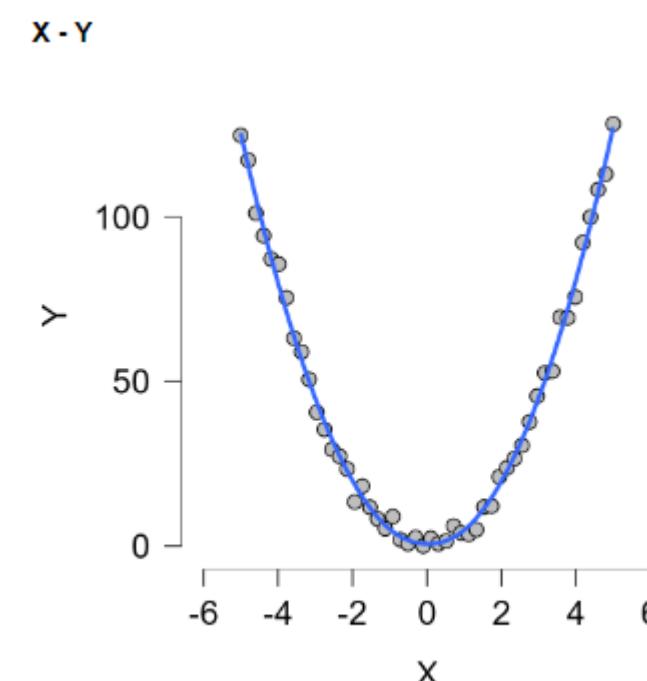
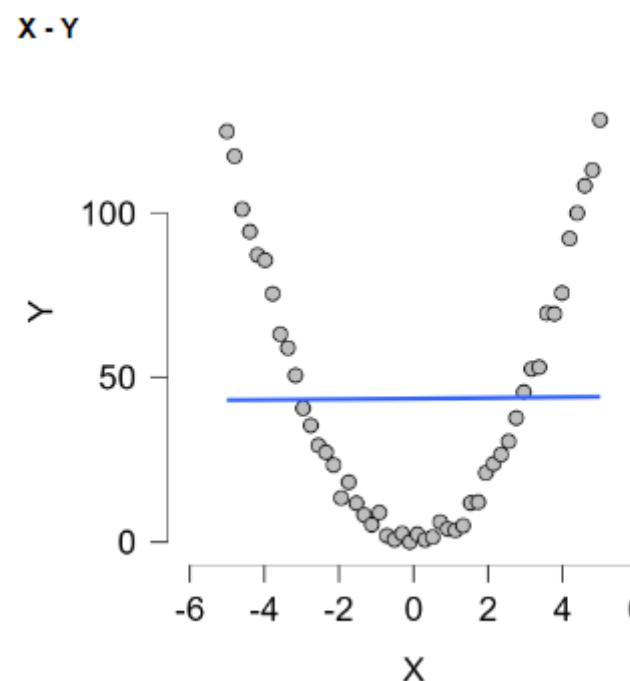


相関の注意点

- 直線的な関係しか意味が無い
 - 単調増加・単調減少
 - 途中で大小に変化があるものは説明できない
- 非線形データや周期性のあるデータは意味が無い
- 極端な外れ値があると相関は下がる
 - そのデータをどうするかは、解析が決めるのではなく研究者が考える

相関係数の意味がない

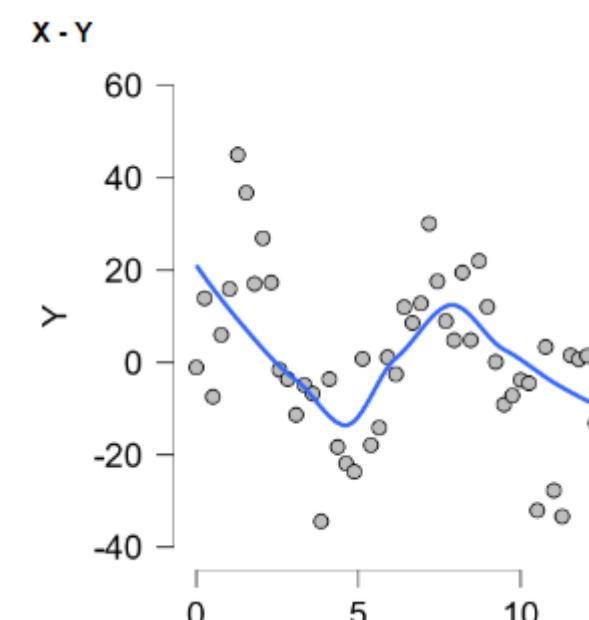
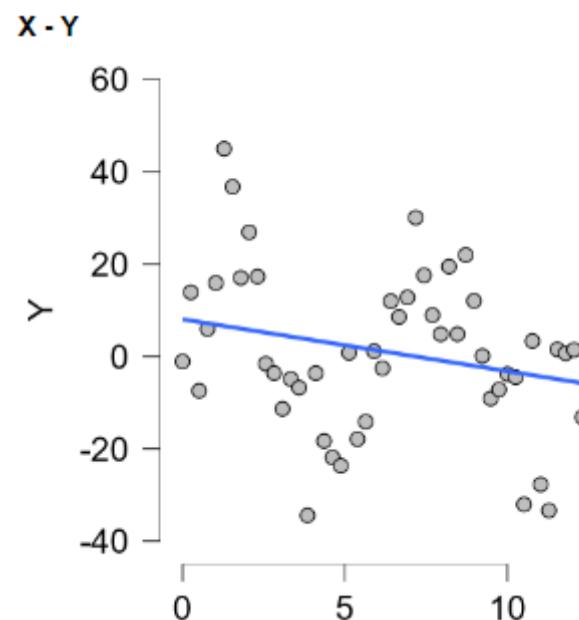
- 2次曲線



ピアソンの 相関			
		ピアソンのr	p
X	-	Y	
		0.007	.961

相関係数の意味が無い

- 周期性データ



ピアソンの 相関

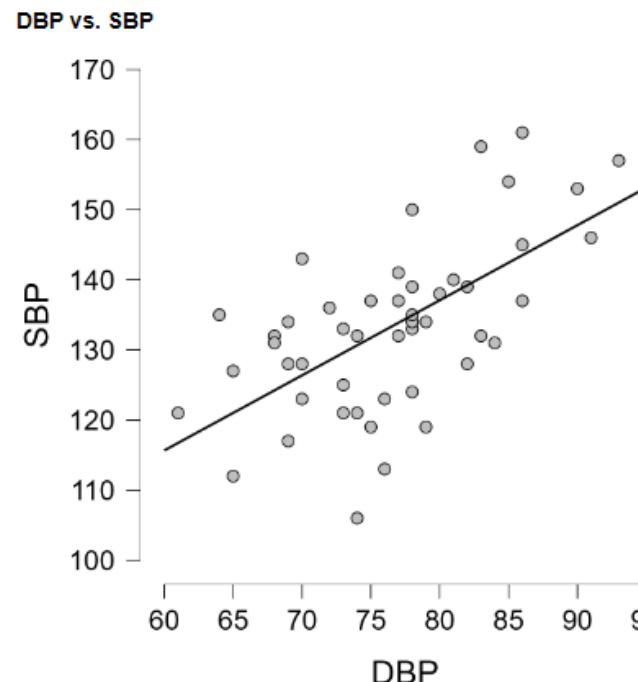
		ピアソンのr	p
X	-	Y	
		-0.243	.089

外れ値1つで相関は変わる

ピアソンの 相関 ▼

		ピアソンのr	p
DBP	-	SBP	0.648 < .001

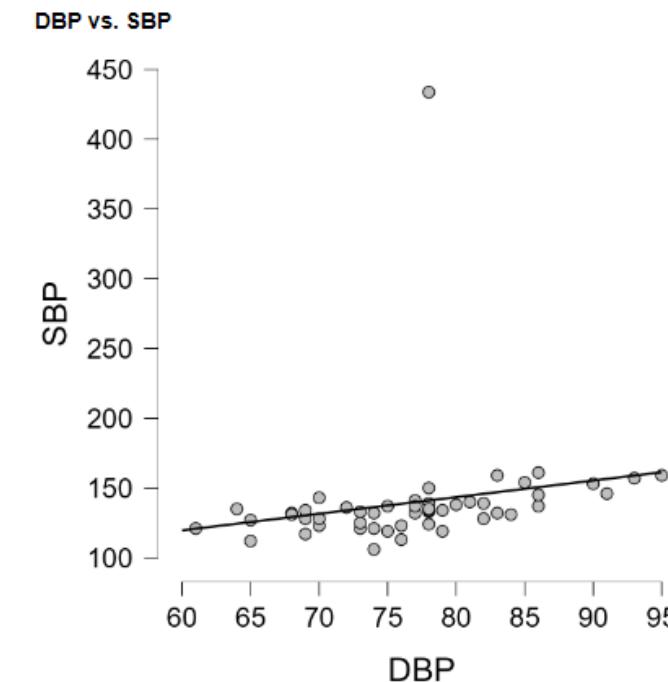
散布図



ピアソンの 相関

		ピアソンのr	p
DBP	-	SBP	0.203 .158

散布図 ▼



相関の注意点

- 因果関係を示しているわけではない
 - どちらが原因で、どちらが結果を示しているわけではない
 - 関連を示しているのみ
- ※因果関係を示すには、原因が結果よりも先に起きていること
が必要(時間的先行性)
- ※因果関係は統計解析ではなく、研究計画や考察で考えること

相関係数の検定

- 相関係数 r の分布は、無相関(0)であるという帰無仮説のもとで、

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

が自由度 $n-2$ のt分布に従う(これがt値)

相関係数の検定とは？

- 相関係数 r は、計算すれば必ず何らかの値が出る
 - 本当は関係がなくても ± 0.2 などになることもある
- 検定で調べていること
 - 得られた r が “偶然の誤差” か？
 - それとも “本当に 0 ではない” のか？
 - r が偶然にしては大きすぎるかを判断
- 相関の検定 = 「 r が 0 でない証拠があるか？」を見ている

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

相関係数の検定

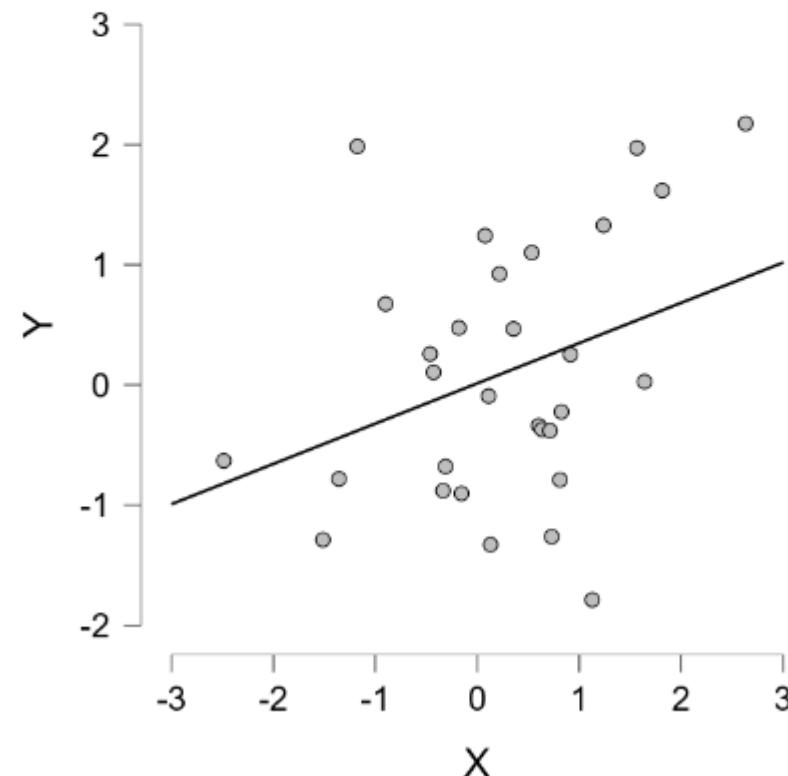
- 相関係数は計算すれば必ず出る
- その値が誤差かどうかは検定が必要
- ただし、データ数が少ないと結果はブレがち

ピアソンの 相関

		ピアソンのr	p
X	- Y	0.337	.069

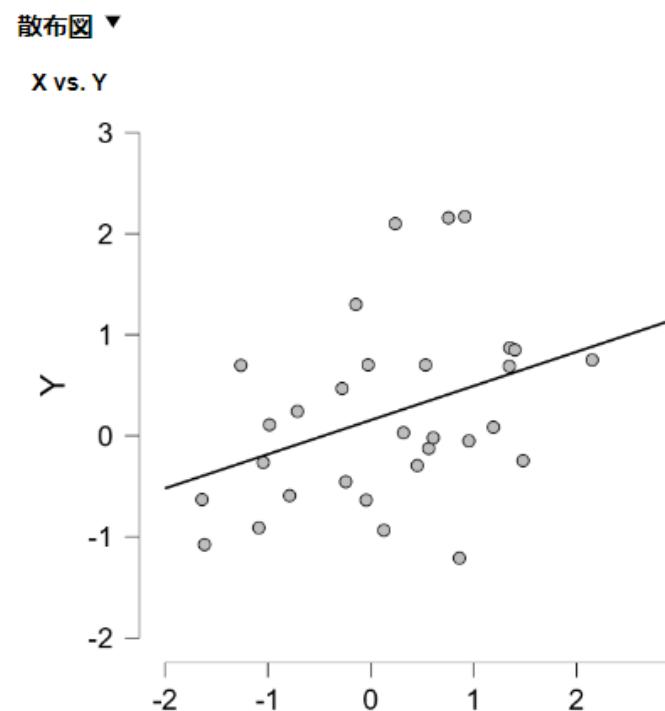
散布図

X vs. Y

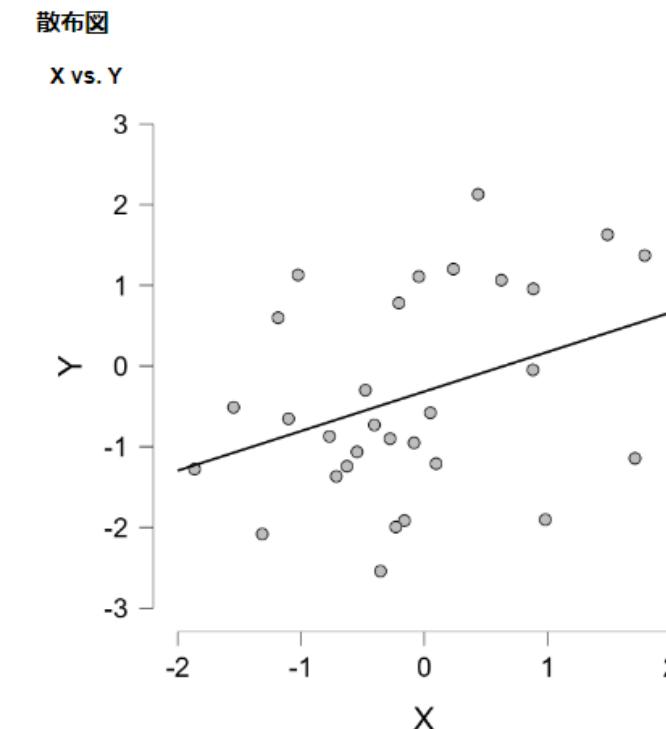


有意確率だけで判断しない

ピアソンの 相関				
		ピアソンのr	p	
X	-	Y	0.368	.046



ピアソンの 相関 ▼				
		ピアソンのr	p	
X	-	Y	0.357	.053



p値は白黒ではなく「濃淡」を示すもの

- データ数(n)が少ないと、p値は大きくぶれる
→ $p=0.049$ と $p=0.051$ の違いは誤差の範囲
- p値は「閾値」ではなく「確率」である
→ p値=0.05 は約束事であり、真実の境界ではない
- p値の本質
→ 帰無仮説が正しいとき、これ以上に極端な結果が出る確率
- 結論は p値だけで決めてはいけない
→ 相関の大きさ、散布図、効果量、信頼区間など
→ 全体の“傾向”として判断する

相関係数の書き方

- と□□は $r=0.\text{xxx}$ ($t=x.\text{xxx}$ d.f.=xx $p=0.\text{xxx}$)
であり、

強い
中程度の
弱い

正の相関
負の相関

がある

無相関である

順位相関

順位相関

- ・ ノンパラメトリックな相関
- ・ データが正規分布近似でなくても大丈夫
 - データの分布を前提としない
 - 順序変数でも問題ない
- ・ 本当は散布図のときに分布を出しておくといい
 - 両方順序変数の時はあまり意味が無い
- ・ 解釈の仕方は、パラメトリックな相関(ピアソンの積率相関)と同じ

順位相関

- Spearman(スピアマン)の順位相関
→ データを順位に変換したピアソンの相関
データは標準化変換している
 d は順位差なので、 d^2 は順位差の2乗
→ リッカート尺度(1~5)などと相性が良い
- Kendall(ケンドール)のて
一致ペアの個数をC、不一致ペアの個数をD
→ ペアの一致・不一致を数える方法で概念が難しい
同順位がある場合は式が複雑に
→ 前後で順序の変化(低ー中ー高)の変化を見るには向いている

$$r = 1 - \frac{6 \sum_i d_i^2}{n^3 - n}$$

$$\tau_a = \frac{C - D}{\frac{n(n - 1)}{2}}$$

相関係数算出の実際

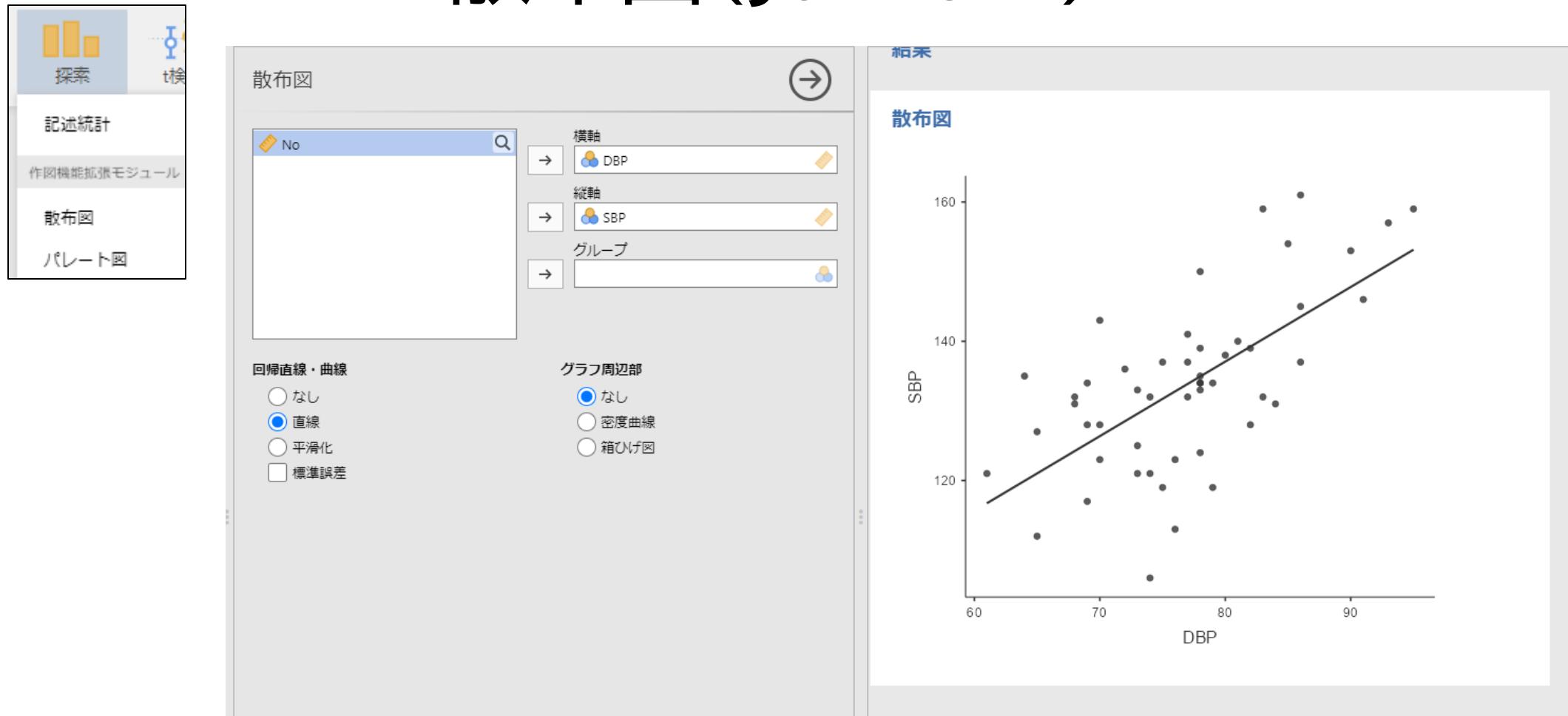
相関係数を出すとき

- 散布図を出力
 - 関係性を確認
 - 外れ値がないかどうか
 - 非線形とか周期データとかは注意
 - 変数が順序変数とわかっている場合は無意味かも
- それから相関係数を算出
- 検定の結果も確認

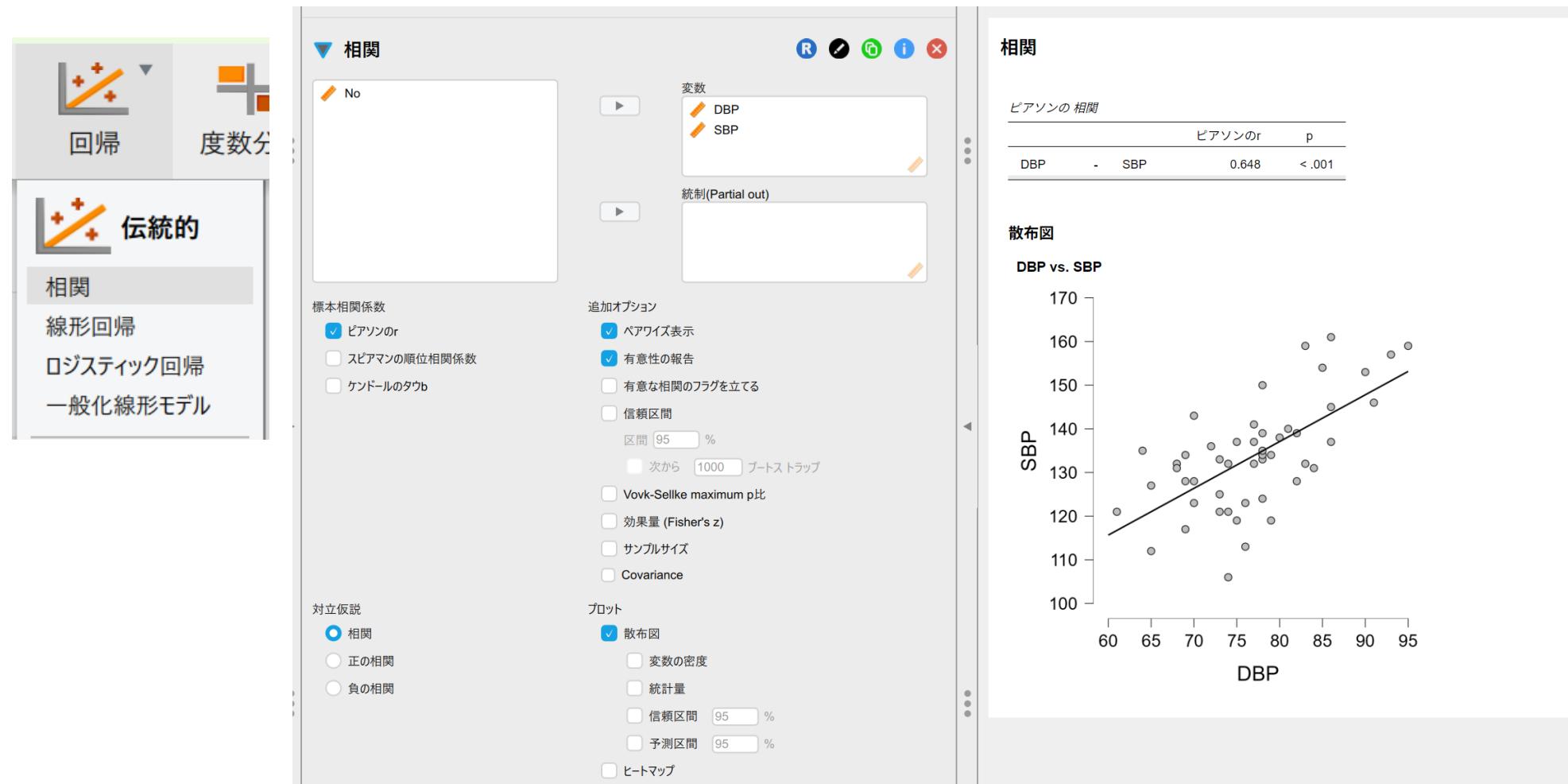
散布図(JASP)



散布図(jamovi)



相関(JASP)





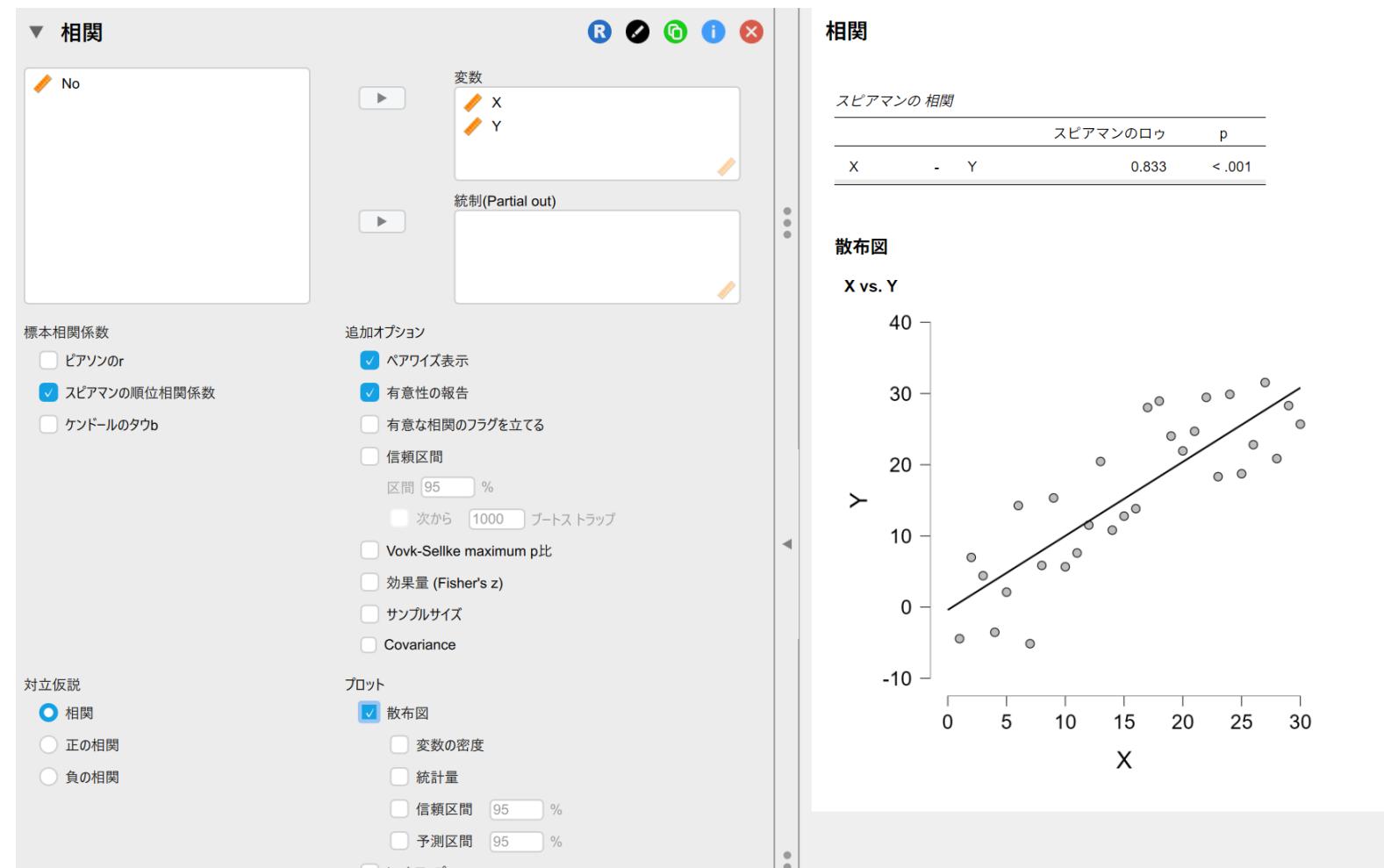
相関(jamovi)

相関行列

	DBP	SBP
DBP	ピアソンの相関係数 自由度 p値	— — —
SBP	ピアソンの相関係数 自由度 p値	0.648 48 < .001

グラフ

順位相関 スピアマン(JASP)



順位相関 ケンドール(JASP)



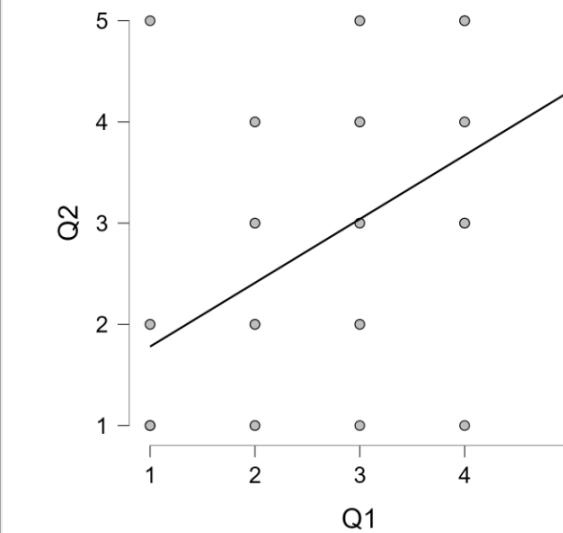
結果

相関

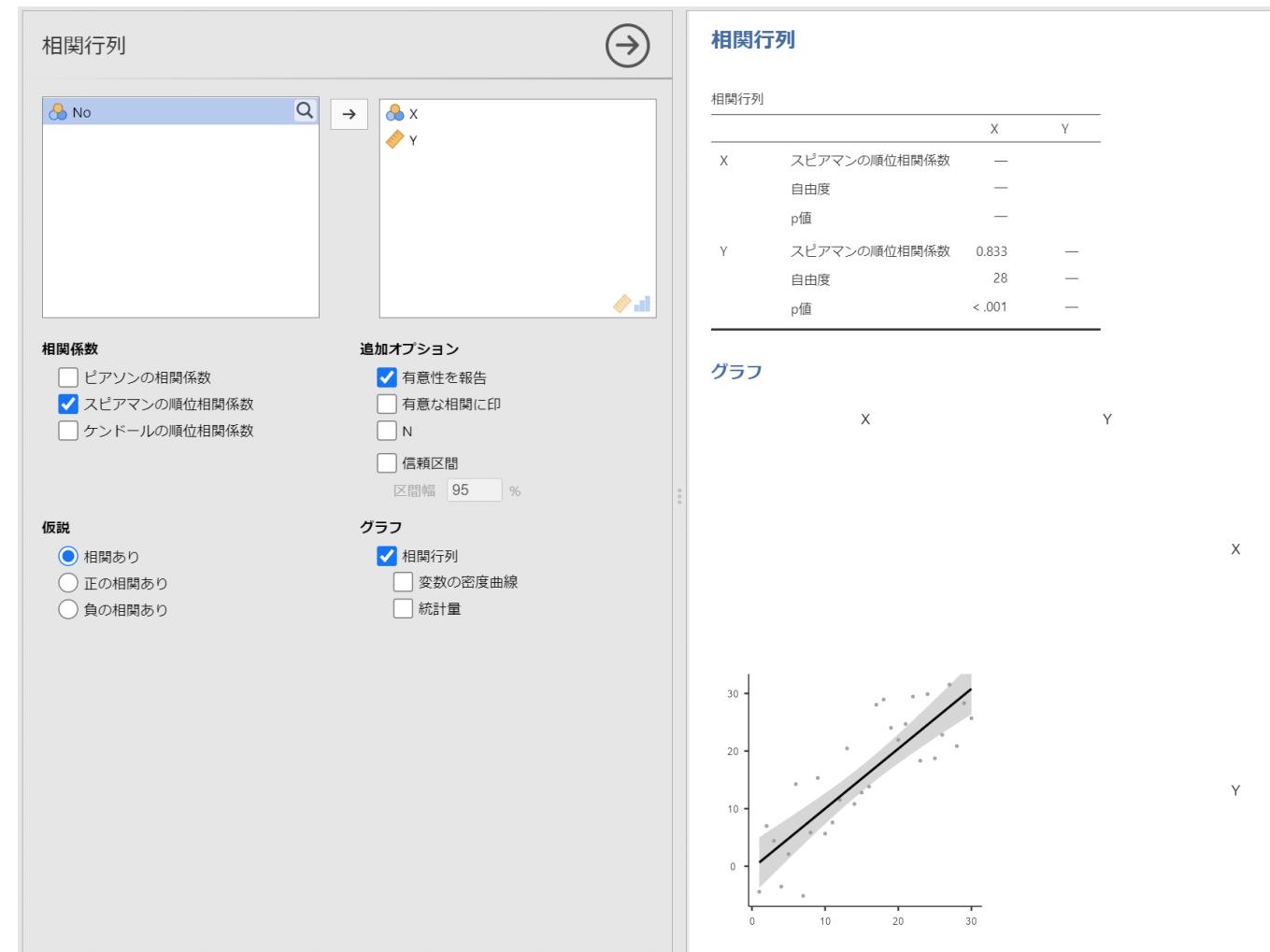
ケンドールのタウ相関

	ケンドールのタウB	p
Q1 - Q2	0.519	< .001

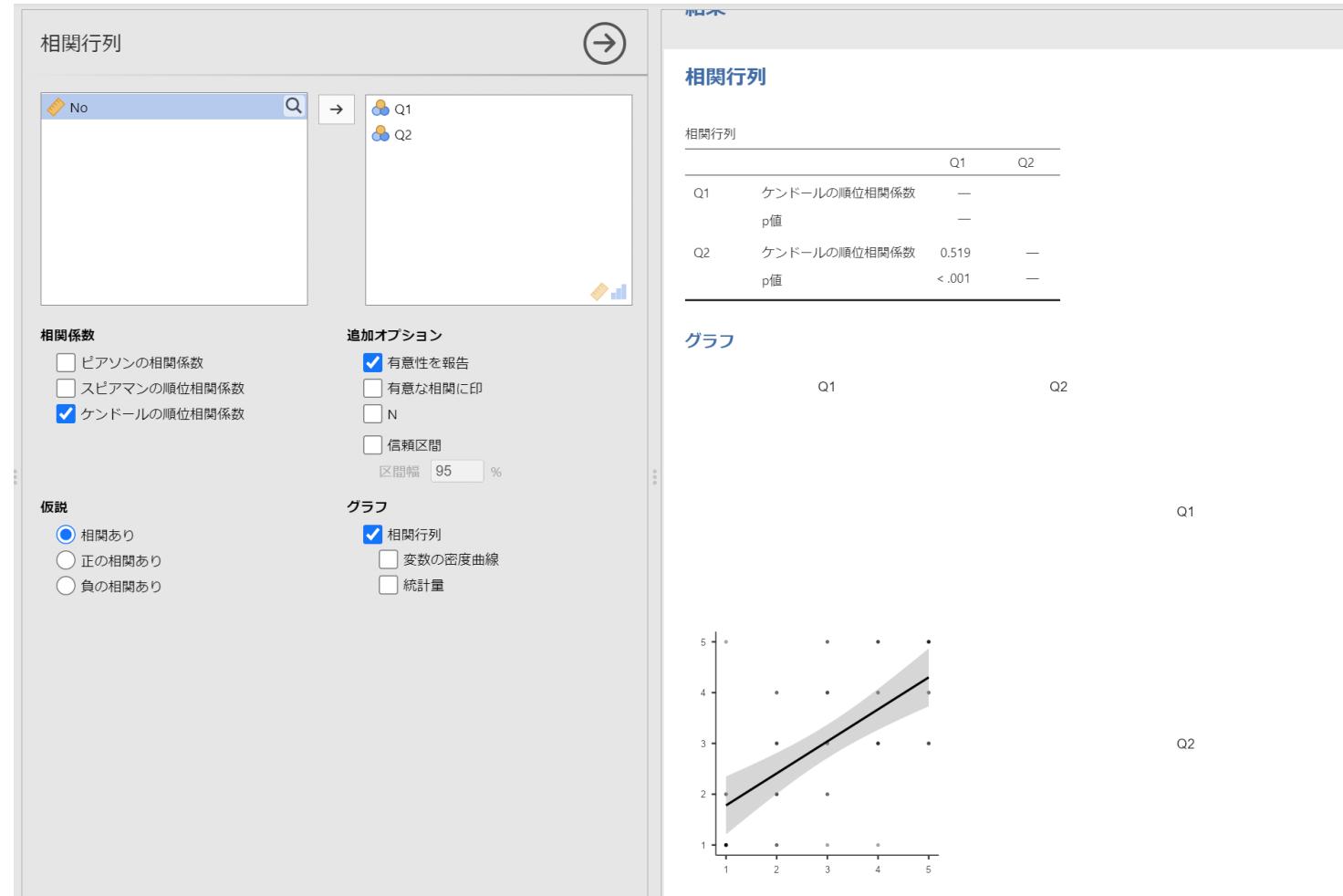
散布図

Q1 vs. Q2

順位相関 スピアマン(jamovi)



順位相関 ケンドール(jamovi)



まとめ

- ・相関係数は線形関係のみ
- ・データの分布状態で計算方法を選ぶ
- ・相関係数は必ず計算されるので、検定結果とセットで確認
- ・ただし、有意確率だけで判断をしない