

IBM SPSS Statistics による 統計解析

【初級編 B】 回帰分析とロジスティック回帰分析



1

相関係数

相関係数 correlation coefficient は、2つの変数の**線形** linear の関係性を評価するための統計量です。相関係数を利用することで、変数間に関係性があるかどうか、関係の方向やその強さを調べることができます。ただし、相関係数は線形の関係性を示すに過ぎず、外れ値の影響を受けるため、**散布図** scatter plot を利用して変数の関係性を視覚的に評価することが大切です。相関係数は、回帰分析や因子分析などの多変量解析の基礎ともなる重要な指標になります。

Contents

- 相関係数の概要
- 散布図
- Pearsonの積率相関係数
- 無相関検定
- Spearmanの順位相関係数
- 偏相関係数

Keyword

相関係数 / 散布図 / Pearsonの積率相関係数 / Spearmanの順位相関係数 / 偏相関係数

§1.1.1 相関係数の概要

相関係数 correlation coefficient は、ある変数 x と別の変数 y の関係性を調べるために用いられる基本的な統計量です。2つの変数の関係性を調べるところから、2変量解析の手法と考えることができます。相関係数は、主に**連続変数(スケール)**や**順序尺度**の変数で適用され、その関係性が数値で要約されますが、単純な線形の指標として要約されてしまうため、以下のような散布図を用いて、変数間の関係性を視覚化して調べることが重要です。

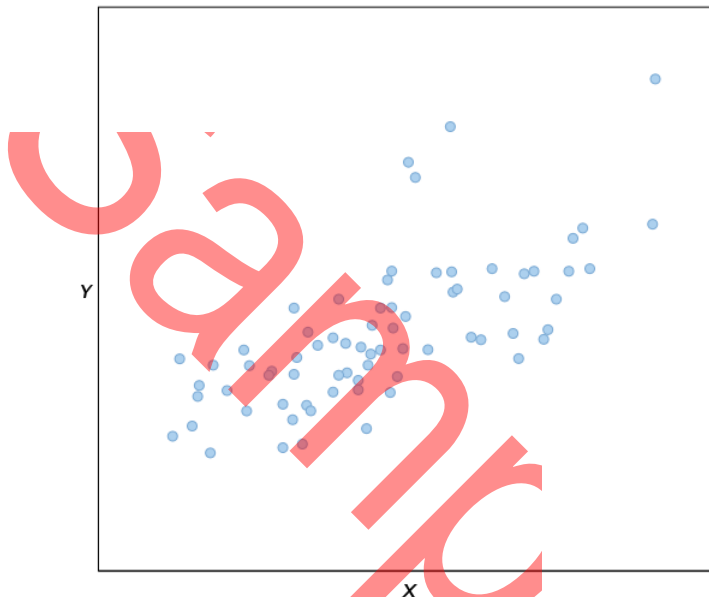


Figure1.1.1 変数 x と変数 y の相関を示す散布図

例えば、身長から体重を説明する場合、身長が x 、体重が y になります。散布図から、身長 x の増加に伴い、体重 y が増加することを視覚的に確認することができます。

TIPS

ある変数 x から別の変数 y を説明するような説明関係を規定する場合は、相関係数ではなく**回帰分析**を用います。

§1.5.1 単純散布図の作成

この例では、身長 **height** と体重 **weight** の関係を分析してみます。相関係数の計算を行う前に、**散布図** scatter plot によって、2つの変数の関係性を調べておきます。

操作手順

1. **グラフメニュー** > **レガシーダイアログ** > **散布図/ドット**を選択します。

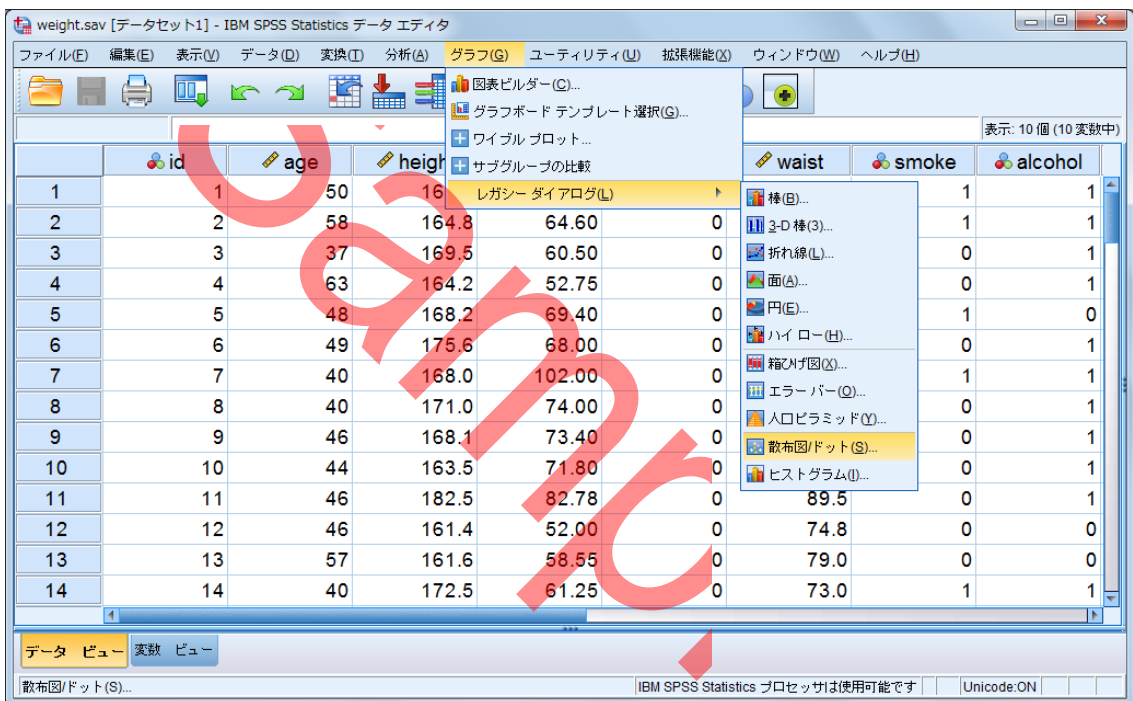


Figure1.5.1 レガシーダイアログからの散布図作成メニュー

TIPS

レガシーダイアログは、変数を選択するほか簡便的な指定で様々なグラフを作成することができます。一方、**図表ビルダー**を利用すると、結果のプレビューをみながらグラフを作成することができます。

操作手順

2. **単純な散布**を選択して、**定義**ボタンをクリックします。
3. **weight** (体重)を**Y軸**ボックスに移動します。
4. **height** (身長)を**X軸**ボックスに移動します。

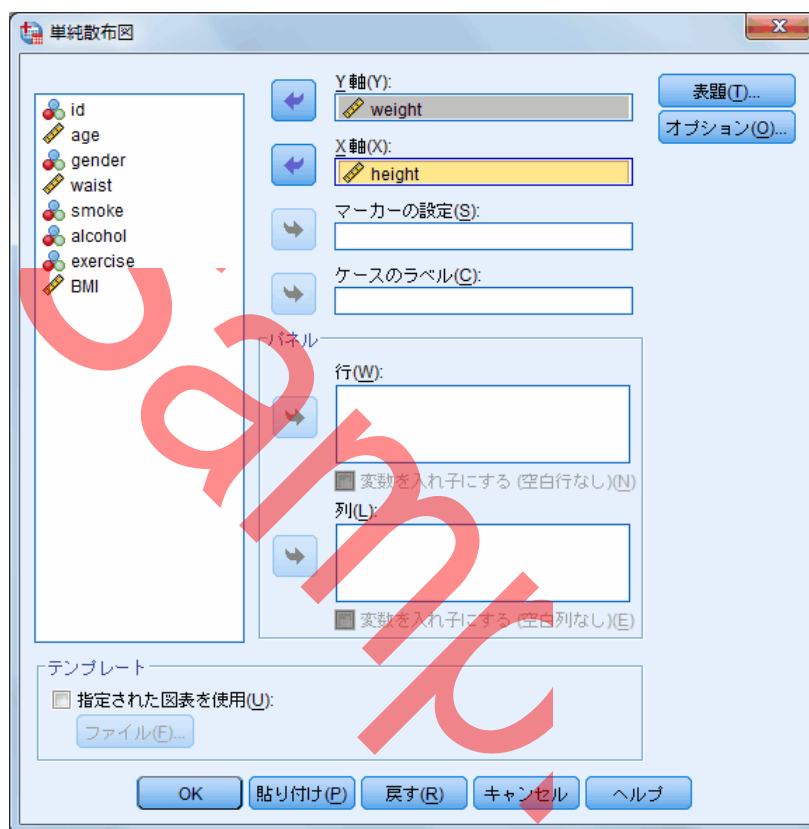


Figure1.5.2 散布図のY軸とX軸の指定

散布図を作成するためには、**Y軸**と**X軸**の指定が必要です。基本的には**連続変数(スケール)**を利用します。変数間に説明関係を仮定する場合は、説明する変数をX軸に置き、説明される変数をY軸に置くのが基本です。

マーカ-の設定を利用することで、散布図のプロットをカテゴリで色分けすることができます。例えば、性別ごとにプロットして、男女の傾向の違いがあるかを評価することができます。また、**パネル**を利用することで、散布図を層別に出力することも可能です。マーカ-やパネルに指定するのは、質的変数(カテゴリ)です。

POINT

散布図のX軸に独立変数、Y軸に従属変数を配置するのが基本です。変数間に具体的な説明関係を仮定しない場合でも、より興味の強いほうをY軸に配置します。

操作手順

5. **OK**ボタンをクリックします。

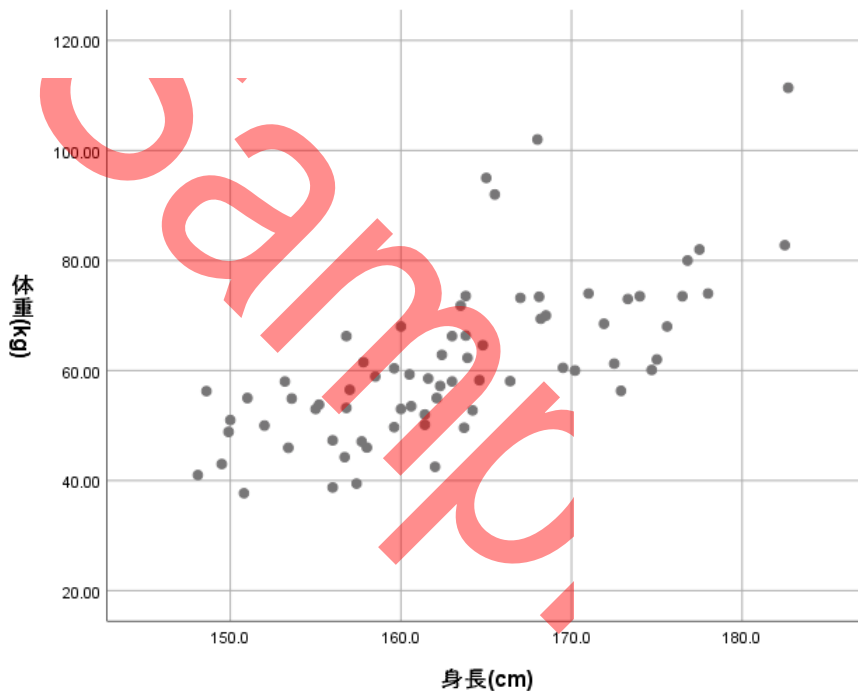


Figure1.5.3 身長と体重の散布図

散布図が出力されます。**Y軸**が体重で**X軸**が身長を表しています。プロットはX軸の値が大きくなるにつれ、Y軸の値も大きくなることを示しています。つまり、身長が高くなるほど体重も重くなる正の関係性があります。

§1.6.1 ピアソンの相関係数の計算

この例では、**weight** (体重)、**height** (身長)、**smoke** (喫煙) の組合せによって、ピアソンの相関係数を計算してみます。同時に3つの変数を利用しますが、相関係数はその2つの組合せごとに計算されます。

操作手順

1. 分析メニュー > 相関 > 2変量を選択します。

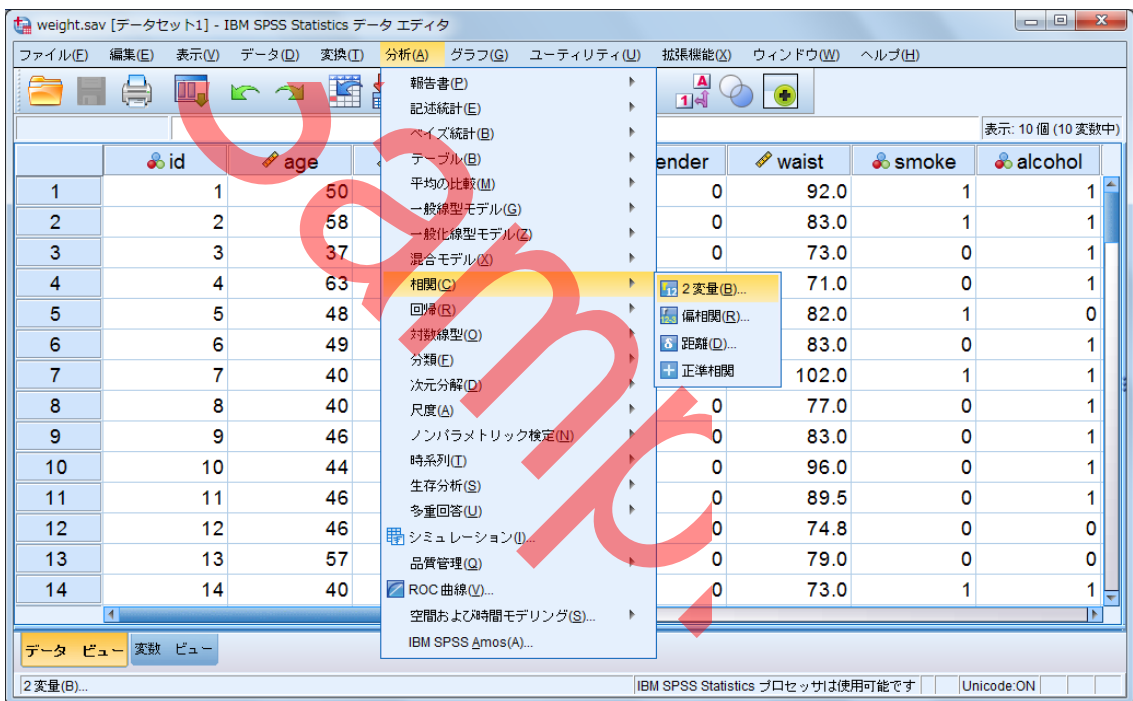


Figure1.6.1 2変量の相関メニュー

操作手順

2. 変数のボックスに、**weight** (体重)、**height** (身長)、**smoke** (喫煙) を移動します。

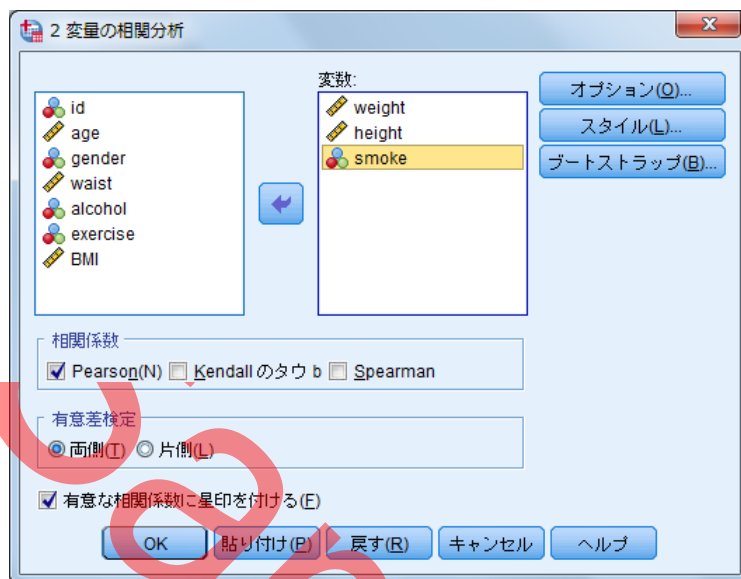


Figure1.6.2 2変量の相関分析ダイアログボックス

変数ボックスに指定するのは、**連続変数(スケール)**です。ただし、等間隔とみなすことができる5段階以上の**順序尺度**を利用することもできます。また、**2値変数**は比率尺度としてそのまま相関係数の計算に利用することが可能です。例えば、**smoke** (喫煙の有無) は名義尺度変数ですが、喫煙なしを0、喫煙ありを1で記録する2値変数であるため、相関係数の計算に用いることができます。相関係数の種類はデフォルトの**Pearson**を利用します。

POINT

ピアソンの相関係数を用いる変数の測定の尺度は、基本的に連続変数(スケール)です。また、等間隔とみなすことができる順序尺度、2値変数も対象とすることができます。

操作手順

3. **OK**ボタンをクリックします。

§1.6.2 ピアソンの相関係数の解釈

相関分析テーブルに、ピアソンの相関係数をはじめとする統計量が出力されます。

		体重(kg)	身長(cm)	喫煙
体重(kg)	Pearson の相関係数	1	.690**	.353**
	有意確率 (両側)		.000	.003
	度数	72	72	71
身長(cm)	Pearson の相関係数	.690**	1	.211
	有意確率 (両側)	.000		.077
	度数	72	72	71
喫煙	Pearson の相関係数	.353**	.211	1
	有意確率 (両側)	.003	.077	
	度数	71	71	71

** 相関係数は 1% 水準で有意 (両側) です。

Figure 1.6.3 ピアソンの相関係数を出力した相関行列

度数は相関係数の計算に用いられたサンプルサイズで、2つの変数の組み合わせごとに計算されます。欠損値が含まれるケースは、ペアごとに除外されます。

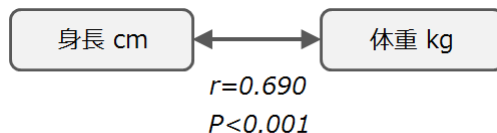
Pearsonの相関係数は、2つの変数間の(線形の)関係の強さをあらわす指標です。その絶対値が1に近いほど関係が強く、0に近いほど関係が弱いことを意味します。

POINT

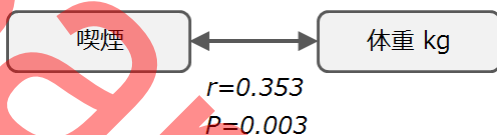
相関係数はその絶対値の大きさによって関係性の強さを解釈することができます。適用分野や目的によって解釈の目安は異なりますが、一般的に以下の目安がよく利用されます。 $|r| \geq 0.7$ の場合に強い相関と解釈されることが多いようです。

$0.0 \leq r \leq 0.2$	ほとんど相関なし
$0.2 \leq r \leq 0.4$	弱い相関が認められる
$0.4 \leq r \leq 0.7$	かなり相関が認められる
$0.7 \leq r \leq 1.0$	強い相関が認められる

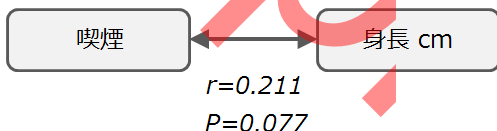
体重と相関が強いのは、**身長**($r = 0.690$)です。正の相関であり身長が高くなると体重が重くなる関係性が認められます。有意確率(両側)からこの相関係数は有意です($P < 0.001$)。



また、**喫煙**と**体重**の間にも正の弱い相関が認められます($r = 0.353$)。これは、正の相関ですので一方の増加に伴いもう一方も増加する関係です。具体的には、喫煙が1増加することで(喫煙なし0→喫煙あり1)、体重が増加することを意味します。つまり、喫煙している人のほうが、体重が重い傾向にあるようです。この相関係数も有意であることが分かります($P = 0.003$)。



なお、**喫煙**と**身長**の間にも正の弱い相関が認められます($r = 0.211$)です。これは、喫煙が1増加することで(喫煙なし0→喫煙あり1)、身長が増加することを意味します。ただし、この相関係数は5%水準で有意ではありません($P = 0.077$)。



POINT

有意確率(両側)の値は、母相関係数が0かどうかの仮説検定の結果となるP値です。帰無仮説「母相関係数は0である」が棄却される場合に、有意な相関係数であると解釈します。5%水準で有意($P < 0.05$)な相関係数については、アスタリスク*が1つ付き、1%水準で有意($P < 0.01$)な相関係数にはアスタリスク**が2つ付きます。

POINT

$P \geq 0.05$ の場合、その相関係数は5%水準で有意ではありません。有意でない相関係数は母集団では無相関である可能性を否定できませんので、その解釈は原則として標本のみにとどめる必要があります。

POINT

$P < 0.05$ の場合、その相関係数は5%水準で有意であると解釈することができます。有意な相関係数は、母集団の真の相関係数が少なくとも0ではないことを意味します。その相関の強さは、標本で求めた相関係数の大きさに基づいて解釈します。

POINT

相関係数は、単純に2つの量的変数の直線的な関係性の強さや方向を示すだけです。他の変数の影響を考慮に入れた分析を行うためには、**偏相関係数**や**重回帰分析**などの分析手法が必要です。