Chapter 1	1 相関	係数 1-1
1.	.1.1	相関係数の概要1-2
1.	.1.2	散布図1-3
1.	.1.3	正の相関、負の相関、無相関1-4
1.	.2.1	ピアソンの相関係数1-7
1.	.2.2	相関係数の検定1-9
1.	.2.3	スピアマンの相関係数1-10
1.	.3.1	疑似相関と偏相関係数1-12
1.	.4.1	使用できる変数の尺度1-14
1.	.4.2	相関係数による分析の実行手順1-16
1.	.4.3	サンプルデータの確認1-17
1.	.4.4	オプションの設定1-19
1.	.5.1	単純散布図の作成1-22
1.	.5.2	単純散布図の編集1-25
1.	.6.1	ピアソンの相関係数の計算1-28
1.	.6.2	ピアソンの相関係数の結果の解釈
1.	.6.3	相関行列テーブルの編集
1.	.7.1	スピアマンの相関係数の計算1-35
1.	.7.2	スピアマンの相関係数の結果の解釈1-37
1.	.8.1	偏相関係数の計算1-39
1.	.8.2	偏相関係数の結果の解釈1-41
Chapter 2	2 回帰	分析 2-1
2.	.1.1	回帰分析の概要2-2
2.	.1.2	回帰式2-4
2.	.1.3	回帰式の説明力2-6
2.	.1.4	残差2-7
2.	.1.5	多重共線性2-8
2.	.1.6	独立変数の自動選択2-11
2.	.1.7	使用できる変数の尺度2-14
2.	.1.8	ダミー変数
2.	.1.9	回帰分析の実行手順2-16

2.2.1	サンプルデータの確認	2-17
2.2.2	オプションの設定	2-19
2.3.1	単回帰分析の実行	2-22
2.3.2	単回帰分析の結果の解釈	2-25
2.3.3	単回帰分析の実行 - 2値の独立変数	2-31
2.3.4	単回帰分析の結果の解釈 - 2値の独立変数	2-33
2.4.1	重回帰分析の実行	2-36
2.4.2	重回帰分析の結果の解釈	2-40
2.4.3	多重共線性の確認	2-48
2.5.1	ステップワイズ法による変数選択	2-52
2.6.1	残差分析 - 影響力と残差の統計量	2-60
2.6.2	残差分析 - グラフによる診断	2-64

Chapter 3	ジスティック回帰分析	3-1
3.1.1	ロジスティック回帰分析の概要	3-2
3.1.2	オッズとオッズ比	3-4
3.1.3	二項ロジスティック回帰	3-6
3.1.4	使用できる変数の尺度	3-7
3.1.5	ダミー変数	3-8
3.1.6	適合度の評価	3-9
3.1.7	多重共線性の問題	3-14
3.1.8	ステップワイズによる変数選択	3-15
3.1.9	ロジスティック回帰分析の手順	3-17
3.2.1	サンプルデータの確認	3-18
3.2.2	オプションの設定	
3.3.1	ロジスティック回帰分析の実行 -単変量-	
3.3.2	ロジスティック回帰分析の結果の解釈 -単変量	3-25
3.4.1	ロジスティック回帰分析の実行 -ダミー変数-	3-29
3.4.2	ロジスティック回帰分析の結果の解釈 -ダミー変数	3-33
3.5.1	ロジスティック回帰分析の実行 -多変量-	3-36
3.5.2	ロジスティック回帰分析の結果の解釈 -多変量	3-41

3.5.3	ロジスティック回帰分析の結果の解釈 -予測精度-	3-45
3.6.1	ROC 曲線 -分析の実行-	3-47
3.6.2	ROC曲線の解釈 -AUC-	3-49
3.6.3	ROC曲線の解釈 ーカットオフ値の探索-	3-51
3.6.4	ROC曲線の解釈 ーカットオフ値の変更ー	3-55
3.7.1	ステップワイズ法による変数選択	3-58
3.7.2	ステップワイズ法による変数選択の結果の確認	3-61

A-1	ダミー変数の作成…	Appendix A
Ę	ダミー変数の概要.	A.1.1
らダミー変数の作成A-3	変数の計算による	A.2.1





本書では、IBM SPSS Statistics for Windows 27 を使用しています。 オプションとして、Regression を使用しています。 IBM およびSPSSは、International Business Machines Corp.の登録商標です。

本書を無断で複写複製(コピー)することは、著作権法上の例外を除き、禁じられています。

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

1



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。

- 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。

🅼 単純散布図		×
 id ✓ age gender 	Y 軸(Y): ✓ weight X 軸(X): ✓ beight	表題(I) オブション(<u>O</u>)
 waist smoke alcohol exercise 	マーカーの設定(<u>S</u>):	
🖋 BMI	「パネル	
	→	
	<u> </u> 列 <u>止</u>	
- テンブレート □指定された図表を使	2000日 2000 2000 2000 2000 2000 2000 200	Ξ)
- 7 7 1 №(⊑) ОК	貼り付け(2) 戻す(2) キャンセル	ヘルプ

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。

マーカー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。マーカー 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。

§1.5.2 単純散布図の編集

次に、変数間の関係性を直線で要約するために、散布図に**回帰直線**を当てはめます。グラフ の編集を行うためには、対象となるグラフをダブルクリックして表示される、グラフ・エディ ターを用います。

操作手順

- 1. 散布図をダブルクリックします。
- 2. グラフ・エディターの要素メニューから合計での線の当てはめを選択します。



Figure1.5.4 グラフ・エディターによる線の当てはめ機能

この操作によって、散布図に回帰直線をあてはめることができます。デフォルトでは、回 帰直線にあわせて回帰式も表示されます。回帰式のラベルは、グラフ・エディターのプロパ ティを利用することで、表示と非表示を切り替えることができます。

あてはめた直線は右にあがっており、2つの変数に正の関係性があることが確認できます。 この直線の具体的な傾きの大きさや切片の値は、回帰分析を実行すると係数として出力する ことができます。



Figure1.5.5 回帰直線を当てはめた散布図(回帰式は非表示)

また、散布図の右上に決定係数**R**²が計算されています。この例では、**R**²=0.476です。これ は変数xによって説明される変数yの分散の割合を表しており、この例では、身長によって体 重の47.6%が説明されると解釈することができます。決定係数については、回帰分析の章で あらためて確認します。

TIPS

グラフ・エディターの要素メニューからデータラベルモードを選択すると、マウスカー ソルが黒く変化し、特定のプロット〇をクリックすることで行番号を表示させることが できます。この操作は外れ値を調べる場合に便利です。

TIPS

グラフ・エディターの任意の場所をダブルクリックするとプロパティウィンドウが表示 されます。プロパティはグラフ・エディターの特定の場所を選択することで、設定でき る内容が切り替わり、軸の編集や色やサイズの変更、特定情報の表示や非表示の設定を 行うことができます。(【入門編A】Chapter5参照)

Sand

§1.7.1 スピアマンの相関係数の計算

スピアマンの相関係数を用いて、weight(体重)、height(身長)、smoke(喫煙)の各組合せの関係性を調べてみます。スピアマンの相関係数はノンパラメトリックの方法です。

操作手順

1. 分析メニュー > 相関 > 2変量を選択します。

ta weight.sav	/ [データセット1] - IBM	SPSS Statistics デー	タエディタ					- 0	×
ファイル(E)	編集(<u>E)</u> 表示(<u>⊻)</u> データ(<u>D</u>)	変換 <mark>(I</mark>)	分析(<u>A</u>) グラフ(<u>G</u>) ユーテ	ィリティ(<u>U</u>)	拡張機能 <mark>(</mark> ,	<u>X)</u> ウィンドウ	(<u>W)</u> ヘルラ	^{\$} (<u>H</u>)
i	🖨 🛄			検定力分析(<u>P)</u> 報告書(<u>P</u>)	>			- 40 /FL (40 T	
		<u>^</u>	<u>^</u>	記述統計 <mark>(E)</mark>	>		衣刀	101回(1038	239X中)
	at de	🖋 age	🖉 heigh	ベイズ統計(<u>B</u>)	>	waist	noke 🖏	n alcoho	4
1	1	50	16	テーブル(B)	· · ·	92.0	1		1 ^
2	2	58	16	平均の比較(M)	,	83.0	1		1
3	3	37	16	→		73.0	0		1
4	4	63	16	→船化線型エデルの	,	71.0	0		1
5	5	48	16	海会エゴル(な)		82.0	1		0
6	6	49	17			02.0	0		1
7	7	40	16			122 2 変量([<u>B)</u> 1		1
8	8	40	17			🛃 偏相関(<u>R)</u> 0		1
9	9	46	16	刈奴線空(<u>U</u>)	ć	る 距離(<u>D</u>)	0		1
10	10	44	16			🛨 正準相關	€ 0		1
11	11	46	18	八元カ府(世)	ĺ.	89.5	0		1
12	12	46	16	べ皮(凶)	· · ·	74.8	0		0
13	13	57	16	ノンハラメドリッン快走(11) 味る利(5)) ~	79.0	0		0
14	14	40	17	時未勿(<u>1</u>)	ĺ.	73.0	1		1 🗸
	<			主任力切(<u>3</u>)					>
データ ビ	ュー 変数 ビュ			∽重回答(<u>U)</u>	>				
				職 シミュレーション(!)					
2 変量(B)				品質管理(Q)	>	能です 🖁	Unicod	e:ON	

Figure1.7.1 2変量の相関メニュー

POINT

スピアマンの相関係数は、順序尺度の変数に用いられる**順位相関係数**ですが、連続変数 (スケール)にも用いることができます。連続変数(スケール)の場合、元のデータ値を昇 順に順位付けした値に対して相関係数が計算されます。

- 2. 変数のボックスに、weight(体重)、height(身長)、smoke(喫煙)を移動します。
- 3. 相関係数の種類としてSpearmanを選択します。
- 4. Pearsonの選択を解除します。
- 5. 下段の三角形のみを表示を選択します。

🏣 2 変量の相関分析			×
	変数(⊻):		オゴション(0)
💑 id	🛷 weight		
🛷 age	s height		スタイル(<u>L</u>)
💑 gender	💑 smoke	•	ブートストラップ(<u>B</u>)
🛷 waist	Sec. 2		
lcohol			
and exercise			
💞 BMI			
有意差検定 ● 両側(I) ○ 片側(L)			
☑有意な相関係数に星印を	titる(E) 🔽 下段の三角形の	みを表示(<u>G</u>) 🗹 対角を表示	Ę
OK	貼り付け(2) 戻す(6	<u>3) キャンセル へ</u>	ルプ
		· · · · ·	
Figure	Ĭ.7.2 スピアマンの	相関係数の出力を	指定

スピアマンの相関係数は、順序尺度の変数や、連続変数(スケール)について正規分布から の大きな逸脱がみられる場合、または外れ値が含まれる場合によく用いられます。また、**下** 段の三角形のみを表示を有効にすると、出力される相関行列の上三角の数字を非表示にして、 相関係数の数値を見やすくすることができます。

POINT

スピアマンの相関係数を用いる場合は、順序尺度の場合、連続変数(スケール)の場合で 正規分布からの大きな逸脱が認められる場合、データに外れ値が含まれる場合などです。

操作手順

6. **OK**ボタンをクリックします。

§1.7.2 スピアマンの相関係数の解釈

Figure 1.7.

相関テーブルに、スピアマンの相関係数をはじめとする統計量が出力されます。このテーブルの解釈方法は、ピアソンの相関係数の場合とほとんど同じです。

			体重(kg)	身長(cm)	喫煙		
Spearmanのロー	体重(kg)	相関係数					
		有意確率 (両側)					
		度数	72				
	身長(cm)	相関係数	.760**				
		有意確率 (両側)	.000				
		度数	72	72			
	喫煙	相關係数	.294	.231			
		有意確率 (両側)	.013	.053			
		度数	71	71	71		
**. 相関係数は 1% 水準で有意 (両側) です。							
*. 相関係数は 5%	水準で有意	(両側) です。					

相関

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。

の相関係数を出力した相関行列

有意確率(両側)は、相関係数が有意かどうかを調べるための無相関検定の結果です。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。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、その解釈は標本のみにとどめる必要があります。

相関係数は、スピアマンの相関係数を表しており、2つの変数間の(線形の)関係の強さを あらわす指標です。

ロジスティック回帰分析の実行 ーダミー変数-§3.4.1

人種 race によって、低体重出生 low にどのような影響があるかを分析してみます。この 例も、従属変数を説明する独立変数(共変量)が1つだけ含まれるモデルになるため、単変量 解析と呼ばれます。人種は、名義尺度変数であるため、ダミー変数に変換する必要があります。 IBM SPSS Statistics のロジスティック回帰分析では、分析実行時にダミーコードに変換する機 能を利用することができます。

従属変数	低体重児出生 low (0なし/1あり)
独立変数(共変量)	人種 race (1白人 /2黒人/3その他)

操作手順

1	. 分析メニ、	ı— > [回帰>二	ニ項ロジステ	ィック を〕	選打				
ta birthwt.sa	v [データセット1] - IBM S	PSS Statistics デー	タエディタ					-		×
ファイル <mark>(E)</mark>	編集 <mark>(E)</mark> 表示(⊻) データ(<u>D</u>)	変換(<u>T</u>) 分析	所(<u>A) グラフ(G)</u>	ユーティリティ	(<u>U</u>)	拡張機能(<u>X</u>)	ウィンドウ(<u>W</u>)	ヘルプ <mark>(</mark> 上	<u>H</u>)
				検定力分析(<u>P)</u> 報告書(<u>P)</u>	T	> >		• Q 表示: 11 f	固 (11 変数	效中)
	al id	🖉 age	# lwt				[⋟] ptl	💑 ht 👘	윩 ui	
1	180	17				(0	0	C) ^
2	15	25		テーフル(旦) 変換の比較(MI)		ĺ.	0	0	1	1
3	89	18				Ś	0	0	1	1
4	93	17		- 動化値刑エデル(<u>□</u>)	7)	Ś	0	0	C	0
5	100	18		渡心球主モリル(2	=/	Ś	0	0	C	0
6	101	18		相関(の)		Ś	0	0	C)
7	102	15		「同県(0)		>			0)
8	113	17		しかり(<u>に)</u> 気が約20円((へ)		· >	目動線型モ:	テリンク… <u>(A</u>)	C)
9	116	17		小坂称 (1)		Ś	👪 線型(<u>L)</u>		C	0
10	117	17		フォ <u>ẹ(に)</u> 、を二(小敏/D)		(🗾 曲線推定(<u>C</u>)		C	0
11	132	18				Ś	🔣 偏相関最小:	2 乗法 <mark>(</mark>)…	1	1
12	133	18		八度(四)	1+&=> /NI)	(🔝 二項ロジスき	ティック <mark>(G</mark>)	1	1
13	143	16		ノンハラメドリック	/19,E(<u>N</u>)	(🛄 多項ロジス 🗄	ティック <mark>(M</mark>)	C	0
14	147	17		時ポッリ(工) 生友公共(の)		ć	᠍ 順序(□)…		C) ~
	<			土1+刀1/1(2) 冬香回答(11)		(P)		>
データ ビ	ュー 変数 ビュー	-	曊	- 多里凹谷(U) シミュレーション(!)		<i>`</i>	▲ 非線型(N)…	_/		
二項ロジステ	ィック(G)			品質管理(<u>Q</u>)		>	🔣 重み付け推算	€(<u>₩</u>)		

Figure3.4.1 二項ロジスティック回帰分析メニュー

- 2. smoke(喫煙の有無)を共変量ボックスから除外します。
- 3. race(人種)を共変量リストボックスに移動します。

🏰 ロジスティック回帰		×
 ♣ id ✔ age ✔ lwt ♣ race ♣ smoke ✔ ptl 	従属変数(<u>D</u>): → low ブロック(B) 1 / 1 前(<u>い</u>) 共変量(<u>C</u>):	カテゴリ(<u>G</u>) 保存(<u>S</u>) オプション(<u>O</u>) スタイル(<u>L</u>) ブートストラップ(<u>T</u>)
 ♣ ht ♣ ui ✔ ftv ✔ bwt 	✓ → → → → → → → → → → → → → → → → → → →	
	ケース選択変数(B) → 「「「」」 のK 貼り付け(P)」 戻す(P)」 キャンセル	ヘルブ
Figure3.4.2 名義	歳尺度を共変量に指定したロジスティック回]帰ダイアログボックス

ロジスティック回帰分析の共変量として利用することができるのは量的変数(連続デー タ)です。2値変数は、間隔尺度とみなすことができるため利用可能です。また、順序尺度は5 段階以上を目安に間隔尺度とみなします。

値を3つ以上含む名義尺度の変数や、3段階または4段階程度の順序尺度は、**ダミー変数**に 置き換える必要があります。また、文字型変数については、値が2つしかないものでもダミー 変数に置き換えなければなりません。(ただし、文字列変数は自動的にカテゴリ共変量ボッ クスに表示されます)。

この例では、<u>人種 race は、白人・黒人・その他の3カテゴリの名義尺度ですので、ダミー</u> コードに変換する必要があります。

- 4. **カテゴリ**ボタンをクリックします。
- 5. raceをカテゴリ共変量ボックスに移動します。

対比の変更では、ダミーコードの方法を指定することができます。ここでは、デフォルトの 指標を利用して、参照カテゴリの設定を確認します。

	×
共変量(<u>C</u>): カテゴリ共変量(<u>T</u>):	_
race(指標)	
<u>*****</u>	
	1
	,
「行り」 キャンセル ヘルブ	
Figure3.4.3 カテゴリ共変量の指定と参照カテゴリの確認	

参照カテゴリ reference category は、カテゴリ変数に関するオッズ比の解釈を行う際の基準となるカテゴリです。解釈の基準としたいカテゴリを選択して、最初 First または最後 Last を指定する必要があります。デフォルトは最後のカテゴリです。

以下の表は、参照カテゴリが最後のその他の場合のコーディングです。<u>race(1)のオッズ比</u> <u>は、その他(0)に対する白人(1)の場合の効果</u>として解釈され、<u>race(2)のオッズ比はその他(0)</u> <u>に対する黒人(1)の場合の効果</u>として解釈されます。

race	race(1)	race(2)
1 白人	1	0
2 黒人	0	1
3その他	0	0

Figure3.4.4 参照カテゴリが「最後」の例

以下の表の参照カテゴリは最初です。つまり、白人が解釈の基準になります。<u>race(1)のオ</u> <u>ッズ比は、白人(0)に対する黒人(1)の場合の効果</u>として解釈され、<u>race(2)のオッズ比は白人(0)</u> に対するその他(1)の場合の効果として解釈されます。



Figure3.4.5 参照カテゴリが「最初」の例

POINT

ダミーコーディングの結果は、ロジスティック回帰分析の結果に出力されますので、オ ッズ比や有意確率の解釈を行う際に、必ず確認するようにしましょう。

操作手順

- 6. 続行ボタンをクリックします。
- 7. OKボタンをクリックします。

以下の表は、変数増加法:尤度比によって方程式に含まれた共変量と、変数減少法:尤度比に よって方程式中に含まれた共変量の違いを示したものです。変数減少法では、人種 raceと喫 煙の有無 smokeが有意な因子としてモデルに含まれています。

変数増加法	尤度比	変数減少法	尤度比
最終月経時体重	lwt	最終月経時体重	lwt
早産の回数	ptl	早産の回数	ptl
		人種	race
	-	喫煙の有無	smoke
高血圧症の既往	ht	高血圧症の既往	ht

Figure3.7.4 変数増加法と変数減少法による方程式中の変数の違い

変数増加法は、従属変数に対して影響のある共変量を1つずつモデルに投入するのに対し、 変数減少法はいったんすべての共変量をモデルに含めてから、従属変数に対する影響のない共 変量を1つずつ減少させていきます。共変量に相関がみられる場合は、変数増加法と変数減少 法でまったく異なるモデルになることは珍しくありません。

POINT

変数増加法だけでなく、変数減少法も試行することが推奨されます。2つのモデルで差異 がみられる場合は、その違いが何に起因するものかを考察します。なお、ステップワイズ 法だけではなく、最終的には強制投入法に戻して、共変量の投入や除去を手動で試行して、 モデルを作成することが重要です。