# 相関分析と回帰分析

目的:2つの分析対象(変数)の間の関係を数値で表す。 (例:中間テストと全国模試の得点の相関係数を算出して、中間テストの点数が 高い生徒が全国模試でも点数が高いかを調査する。)

- ・結果の数値は  $-1 \leq r \leq 1$  (rは相関係数を表す)の範囲になる。
- ・結果の解釈の基準は研究対象によって違いがあるが、一般的には以下のように判断することが多い。
   .00 ~ ± .20 ほとんど相関がない(.00は無相関)
   .20 ~ ± .40 低い(弱い)相関がある
   .40 ~ ± .70 かなり(比較的強い)相関がある
   .70 ~ ± 1.00 高い(強い)相関がある
   ※質問紙を使った研究では.30~.50である程度の相関があると考えられる(Dörnyei, 2001, p.224)
- ・相関係数を表す直線(回帰直線; regression line)を利用して、回帰分析(regression analysis)を 行うことが可能。

## 1. EXCEL を使って分析する

#### 1.1. 相関係数の計算

セルに=correl(テスト1の点数が入力されている列,テスト2の点数が入力されている列)を入力。

	A	В	С	D	E	F	G	Н	I	
1	氏名	性別(1=女	クラス	学籍番号	TOEIC_total	Level2	Level3	Level4	Total	
83	大塚愛	1	3	200784	715	20	19	12	51	
84	矢井田瞳	1	3	200785	605	17	12	11	40	
85	深津絵里	1	3	200786	470	18	13	13	44	
86	倖田來未	1	3	200787	580	19	15	11	45	
87	加藤ローサ	1	3	200788	490	19	16	18	53	
88	押切もえ	1	3	200789	570	18	14	14	46	
89	戸田菜穂	1	3	200790	465	16	9	17	42	
90	川原亜矢子	1	3	200791	510	18	12	14	44	
91	竹中直人	2	3	200792	610	20	14	12	46	
92	三村マサカズ	2	3	200793	380	15	5	9	29	
93	木村佳乃	1	3	200794	440	15	12	11	38	
94										
95										
96					=CORREL(	E2:E93,I2:I9	3)			
97					CORREL	配列1.配列2)	]			
98										

この例では相関係数は 0.64 になっている。

結果の報告を論文に書くときには、「TOEIC の得点と語彙サイズテストの得点の相関を求めたところ、 r = .64 で比較的強い相関があることがわかった。」というような記述方法になる。

#### 1.2. 散布図を描く

「挿入」⇒「グラフ」を選択。

「グラフの種類」の中の「散布図」を選択し、「次へ」をクリック。

グラフ ウィザード - 1/4 - グ	ブラフの種類 🥐 🔀
標準 ユーザー設定	
グラフの種類( <u>C</u> ):	形式①
■ 1/14	· · · ·
の日	
in the interview	
◎ どう () ● 等高線 () ●	
	散布図 - 値の組を比較します。
	サンプルを表示する①
	〕 < 戻♪ _ 沐へ№> 完 <b>7</b> (E)

「データ範囲」に=テスト1の点数が入力されている列,テスト2の点数が入力されている列となるように、データを選択する。



完成。(x軸やy軸に軸ラベルを挿入したければ、グラフ上で右クリックする)



1.3. 回帰分析をしてみる

・ 語彙サイズテストの点数から、TOEICの点数の予測する数式 y = ax + b を作ってみる。

この場合,予測したいテスト(TOEIC)を y 軸にする必要があるので, もう一度, 2.1.2. 散布図を描くのはじめに戻って,「挿入」⇒「グラフ」を選択。

「グラフの種類」の中の「散布図」を選択し、「次へ」をクリック。

「系列」タブでXの値を
語彙サイズテストの列,Yの値を
TOEICの
点数の列となるように列を選択する。

<b>元のデータ</b> ? 🗙
データ範囲 系列
800         ・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・
0 10 20 30 40 50 50
系列(S) 系列(T) 系列(T) スの値(S):=まとあたもの(\$\$\$2\$\$\$93) マの値(C):=まとあたもの(\$\$\$2\$E\$93) マの値(C):=まとあたもの(\$\$\$2\$E\$93) 、
キャンセル 〈戻る(B) 次へ(M) 〉 完了(F)

ツールバーの「グラフ」の中から「近似曲線の追加」を選択。「線形近似」が選ばれているのを確認し て、「オプション」タブを選択。

捐	) 方書 仰人間	Q) ツール(T)	グラフ(C) ウィンドウ(W)	All	プ(H)			
1	<u>u</u>   📰 🗄		グラフの種類(ど)		🖏 - <u>A</u>	- 🚆 🖬 🖬	🗃 🔒 🍙 l	🖪 🛕 🗳
		▽校閲結果の返	元のデータ( <u>S</u> )		(N)			
-	• ·		グラフのオブション(単					
	С	D	近似曲線の追加化		G	Н	I	J
=女	クラス	学籍番号	*		evel3	Level4	Total	
2	1	200703	500	18	13	13	44	
1	1	_			-			
2	1	800						
1	1	700					•	
1	1						. • .	
2	1	600					<del></del>	_
2	1	500				• •	<b>.</b>	
2	1	000			•			
1		- H 400					<u></u>	
2	1	- • <sup>-</sup>			• •			
2	1	300			*** *			
2	1	200			· · ·	·		
1	. 1				•			
2	1	100						
1	1							
2	1		0 10	20	30	40	50	60
1	1				 吾乗サイズテ	자	-	
2	1							
1	1	200722	420	18	14	12	44	

近似曲線の追加	×
種類 オプション	
近似または回帰の種類	
·····································	次數(D): 2 ◆ 項式近似(P)
累乗近似∞ 指数近似∞ 移	区間(E): 2 章 動平均(M)
Total	

「オプション」の「グラフに数式を表示する」と「グラフに R-2 乗値を表示」をチェック

近似曲線の追加	
・         ・         ・	
	OK キャンセル

完成。



表の中の数式, y=10.888x - 13.437 と R<sup>2</sup>=0.4154 は,

### **TOEIC**の予想点 = (10.888 × 語彙サイズテストの点数) - 13.437 という回帰式が約42パーセントの説明率を持っていることを意味している。

 $r^{2}$ (決定係数,説明率)は見たとおり、Step 1 相関係数の計算で求めた相関係数 (0.64)の二乗になっている。

# 2. SPSS を使って分析する

2.1. 相関係数の計算

「分析」⇒「相関」⇒「2変量」を選ぶ。

まと	め.sav [データセ	<u>&gt;</u> 1]	- SPSS	; <del>7*-9</del> I	j*19					
ファイル(	E) 編集(E) 表示	F.∭	データ( <u>D</u> )	変換(T)	分析( <u>A</u> ) ゲラフ( <u>G</u> ) ユ	ーティ	Jティ( <u>U</u> )	ウィンドウ	( <u>₩)</u> ^//フ°(	( <u>H</u> )
<b>&gt;</b>	1 🔒 🖬 🔸	<del>ب</del> ا	- <b>[</b> ?	M 📲 I	報告書(P) 記述統計(E)	;				
1:氏4	2		島田紳	助	平均の比較(M)	•				
	氏名	性別	クラス	学籍番号	→ 股線型モデル( <u>G</u> ) 温会エニショムム	-1	el3	Level4	VocTotal	Vě
1	島田紳助	2	1	200703	混合て7 IV 🖤	- 1	10	12	44	
2	長澤まさみ	1	1	200704	相関(C)	_	- 2 溪口	■( <u>B)</u>	32	
3	渡辺謙	2	1	200705	回帰( <u>R</u> )	•	偏相	関( <u>R</u> )	37	
4	松嶋菜々子	1	1	200706	対数線型(Q)	•	距離	( <u>D</u> )	44	
5	天海祐希	1	1	200707	分類(Y)	- 11	14	8	40	
6	桑田佳祐	2	1	200708	データの分解( <u>D</u> )	- <b>F</b>	14	12	44	
7	稲垣吾郎	2	1	200709	尺度( <u>A</u> )	- <b>F</b>	14	13	45	
8	郷ひろみ	2	1	200710	ノンパラメトリック検定(N)	- <b>F</b>	11	8	36	
9	水野真紀	1	1	200711	生存分析(S)	-+	10	9	37	
10	宮崎あおい	1	1	200712			13	12	42	
11	浜田雅功	2	1	200713			13	9	38	
12	ビートたけし	2	1	200714	<u>A</u> mos 6		16	14	47	
13	大久保嘉人	2	1	200715	260 9	_	9	8	26	

左のリストから相関係数を計算したい変数を選んで右側に移したら、「OK」をクリック。

■ 2 変量の相関分析	X
<ul> <li>◆ 性知!(1=女, 2=男) [性別!(</li> <li>◆ クラス</li> <li>◆ 学籍番号</li> <li>◆ Level2</li> <li>◆ Level4</li> </ul>	OK 貼り付け(P) 戻す(R) キャンセル ヘルフ°
相関係数 ✓ Pearson(N) 【Kendall の対 b(K) 【Spearman(S) 本音美格学	_
(月過至1000年 (○両側①) ○片側①	
✓ 有意な相関係数に星印を付ける(E) 75%ョン(Q)	

結果がアウトプットされる。

		TOEIC	VocTotal
TOEIC	Pearson の相関係数	1	.645(**)
	有意確率(両側)		.000
	Ν	92	92
VocTotal	Pearson の相関係数	.645(**)	1
	有意確率(両側)	.000	
	Ν	92	92

### 2.2. 散布図を描く

「グラフ」⇒「散布図/ドット」を選び、「単純な散布図」⇒「定義」。

11 まとめ.sav [データセット1] - SPSS データエディタ											
ファイル(	E) 編集(E) 表示	ΞŴ	データ( <u>D</u> )	変換( <u>T</u> )	分析( <u>A</u> )	かうフ( <u>G</u> ) ユーティリティ(L	D 97				
6	🗄 🖻 🔶	÷ 1	<b>⊨</b> [?	M 📲 I	ĥ 🖽 🤅	図表ビルダー( <u>C</u> ) インタラクティブ <sup>(</sup> <u>A</u> )					
1:氏名	2		島田綿	助		棒(B)					
	氏名	性別	クラス	学籍番号	TOEIC	3-D 椿(3)					
1	島田紳助	2	1	200703	500	折れ線(L)	[				
2	長澤まさみ	1	1	200704	300	面(M)					
3	渡辺謙	2	1	200705	365	円(E)					
4	松嶋菜々子	1	1	200706	320	Λ/ Π−(H)					
5	天海祐希	1	1	200707	365						
6	桑田住祐	2	1	200708	445	パレート図(B)					
	植理音郎	2	1	200709	430	管理区					
8	娜ひろみ	2	1	200710	345	結びげ図へへ					
9	「水野長紀」	1	- 1	200711	330						
10	宮崎のあい	1		200712	410						
10	「渋田雅切」	2	- 1	200713	310	XUC/39NU/					
12	に上げたりし	2	1	200714	410	散布図/ドット(S)	-				
14	] 시시 18 茄八   초미博고	2	- 1	200710	400	ヒストグラムの					
15	伊生院光	2	1	200717	270	正規 P-P プロット(P).					
16	宇多田ヒカル	1	1	200718	415	正規 Q-Q プロット(Q)	)				
17	小田和正	2	1	200719	435	時系列(U)					
18	松田聖子	1	1	200720	385	ROC 曲線(V)					
19	木村拓哉	2	1	200721	325	その他の時系列(T)	- +				
20	田中麗奈	1	1	200722	420	181 14					

散布図/ドット			X
単純な 数布図 オーバーレイ 散布図	行列 載布図 3-D 載布図	シンプル 通 ドット	定義 キャンセル ヘルフ*

Y軸,X軸に変数を選んで移動させて、「OK」をクリック。





完成。

2.3. 回帰分析をしてみる

「分析」⇒「回帰」⇒「線型」を選ぶ。

品 まと	め.sav 「データセ	ንト1]	- SPSS	5 <del>7*-9</del> I	j°19				
ファイル(	E) 編集(E) 表示	÷ 🕐	データ( <u>D</u> )	変換(T)	分析( <u>A</u> ) ゲラフ( <u>G</u> )	ユーティ	リティ(山)	ウィントウ 😡	) ^//フ°(!
	A 🖬 🔶	<del>ب</del> ا	- <b>[</b> ?	M • 🗐 I	報告書(P) 記述統計(E)	•			
1:氏4	2		島田紳	助	平均の比較(M)	•			
	氏名	性別	クラス	学籍番号	一般線型モデル( <u>G</u> ) 温会エニショクク		el3	Level4 Vo	cTotal
1	島田紳助	2	1	200703	記合セアル区		13	13	44
2	長澤まさみ	1	1	200704	相関心		0	7	32
3	渡辺謙	2	1	200705	回)\$\$(E)		線型(	_)	37
4	松嶋菜々子	1	1	200706	对数線型( <u>O</u> )	•	田線	匯定(℃)	44
5	天海祐希	1	1	200707	分類( <u>Y</u> )	•	順 定(	n)	40
6	桑田佳祐	2	1	200708	データの分解( <u>D</u> )	. ▶.	7087 P	<u>v</u>	44
7	稲垣吾郎	2	1	200709	尺度( <u>A</u> )	•	最適	R度法(0)…	45
8	郷ひろみ	2	1	200710	ノンパラメリック検定()	୬≯		0	36
9	水野真紀	1	1	200711	生存分析(S)	•	10	9	37
10	宮崎あおい	1	1	200712	多重回答(U)	•	13	12	42
11	浜田雅功	2	1	200713			13	9	38
12	ビートたけし	2	1	200714	<u>A</u> mos 6		16	14	47
13	大久保嘉人	2	1	200715	260 9		9	8	26

「従属変数」に TOEIC の点数を入れて、「独立変数」に VocTotal を入れて、「OK をクリック」。



## 3. 3 つ以上の変数の相関と重回帰分析

TOEIC の点数と語彙サイズテストだけではなく,動機づけ(motivation)も加えてそれぞれの相関を チェックする方法は, Step 1 相関係数の計算 の「分析」⇒「相関」⇒「2 変量」までは同じで,変数 に motivation も入れて OK をクリックする。

相関係数							
		TO EIC_total	VocTotal	Motivation			
TO EIC_total	Pearsonの相関係数	1	.645**	.294**			
	有意確率(両側)		.000	.004			
	N	92	92	92			
VocTotal	Pearson の相関係数	.645**	1	.275**			
	有意確率 (両側)	.000		.008			
	N	92	92	92			
Motivation	Pearson の相関係数	.294**	.275**	1			
	有意確率 (両側)	.004	.008				
	N	92	92	92			
**. 相関係数は1%水準で有意(両側)です。							

上の結果をエクセルで加工した後の表。

	TOEIC	語彙サイズ	Motivation
TOEIC	-		
語彙サイズ	.65**	-	
Motivation	.29**	.28**	-
N = 92, ** p < .01			

投入済み変動または除去された変動 除去され た変数 投入済み変数 方法 Motivation, VocTotal 投入 a. 必要な変数がすべて投入されました。 b. 従属変数: TO EIC total 王子油集計 推定値 モデジル 82番 R2 垂 a. 予測値: (定数, Motivation, VocTotal。 分散分析 F値 33.589 平方和 450243.140 モデル 自由度 <u>有意確率</u> | 平均平方 미명 残差 596503.327 89 6702.285 全体 1046746.5 91 a. 予測値: (定数)、Motivation, VocTotal。 b. 従属変数: TO EIC total 係翻 標準化係 非標準化係数 右音磁率 B 標準 67.671 .86 -1.027 (定新) 307 VocTotal 10.302 1.406 .610 7.328 .000 1.515 .133 Motivation 19.910 13,144 126 a. 従属変数: TO EIC\_total

結果のアウトプット。

この結果画面上で右クリックして, 「エクスポート」 「ファイルの種類」を 「Excel ファイル (\*.xls)」にして 保存する。

> この表のことを 「相関係数行列表」と呼ぶ。

右上半分(もしくは左下半分)は同じ結果 なので,省略する。 -の部分は1と書いてもよい。

語彙サイズテストと動機づけのアンケート結果を用 いて,TOEIC の点数を予測する重回帰分析を行って みた結果が左のアウトプットになる。

 $r^{2}$ (決定係数,説明率) = .43 となり,語彙サイズテ ストのみでの回帰分析よりも,.01 だけ高い結果にな っている。「調整済み R2 乗」は独立変数の数が複数 あるときに使う。

この有意確率が.05 未満であれば、この回帰式は使えると判断する。

 $y = 10.302x_1 + 19.910 x_2 - 67.671$ ( $x_1 = 語彙サイズテスト, x_2 = motivation$ ) という回帰式が求められた。

しかし,この部分の有意確率が.05以上(.133)なので,「motivationはこの重回帰分析では予測に役立たない」と判断される。

# 4. 相関係数の種類

- ・ピアソンの積率相関係数(Peason's product moment correlation coefficient)
- ・スピアマンの順位相関係数(Spearman's rank correlation coefficient)
- ケンドールの順位相関係数(Kendall's rank correlation coefficient)
   ※ケンドールの順位相関係数はデータが少なくて、同率順位が多いときに使う

各種統計解析ソフトのデフォルトは「ピアソンの積率相関係数」。

<ul> <li>2 支量の相関分析</li> <li>注別(1=女, 2=男) [性別(2 クラス</li> <li>学祥番号</li> <li>グ Level2</li> <li>グ Level3</li> <li>グ Level4</li> </ul>	変数 <u>\(</u>	): DEIC_total perTotal ptivation	ОК           ВЫ(11) Ф)           ЭТ (В)           Тоб (В)           Тоб (В)           Тоб (В)
相関係数 IV Pearson( <u>N</u> ) IV Ken	dall のか b( <u>K</u> )	Spearman(S)	
有息差快定 ⓒ両側①	◯片側Ϣ		
☑有意な相関係数に星印を	付ける(E)	<u>オフジョン(Q)</u>	

順位相関係数を計算したい場合には、SPSS はここにチェックを入れる。

Excel の場合には,関数で「rank」を使い変換してから,ピアソンの積率相関係数を求める方法が 考えられる。この場合,スピアマンの順位相関係数と同じ結果になる。