

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	B	C	D	E	F	G	H	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	俣田来未	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:I93)				
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。

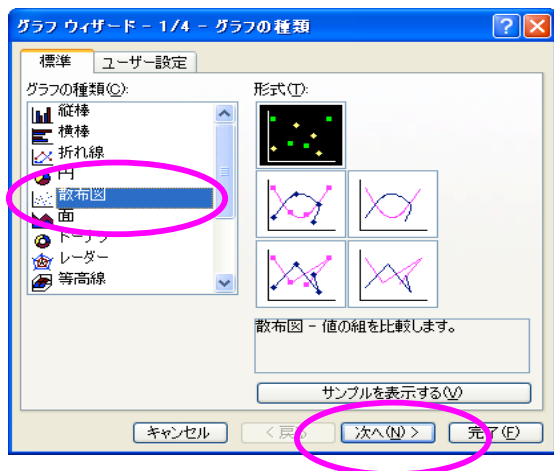
結果の報告を論文に書くときには、「TOEIC の得点と語彙サイズテストの得点の相関を求めたところ、 $r = .64$ で比較的強い相関があることがわかった。」というような記述方法になる。

説明使用データ <http://www.mizumot.com/stats/ch9-10.xls>

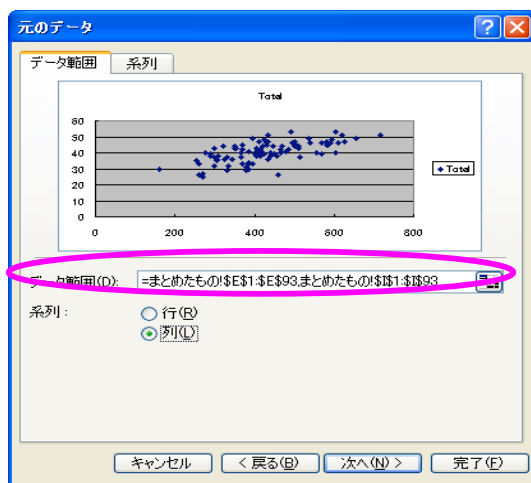
1.2. 散布図を描く

「挿入」⇒「グラフ」を選択。

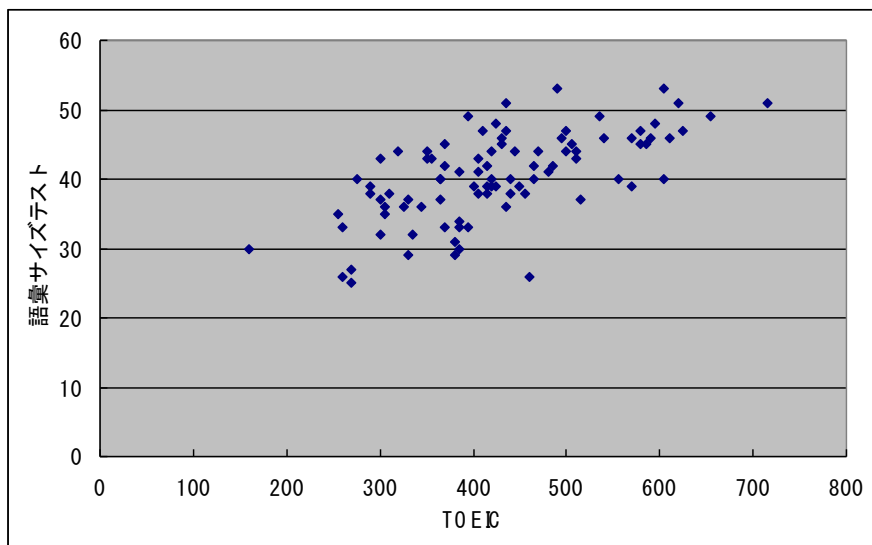
「グラフの種類」の中の「散布図」を選択し、「次へ」をクリック。



「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)



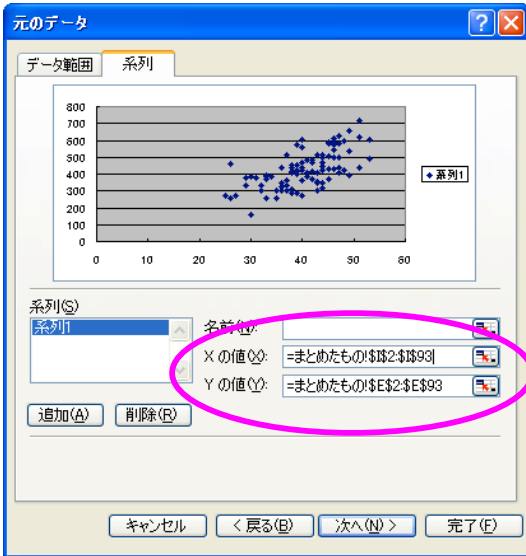
1.3. 回帰分析を試みる

- ・ 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。

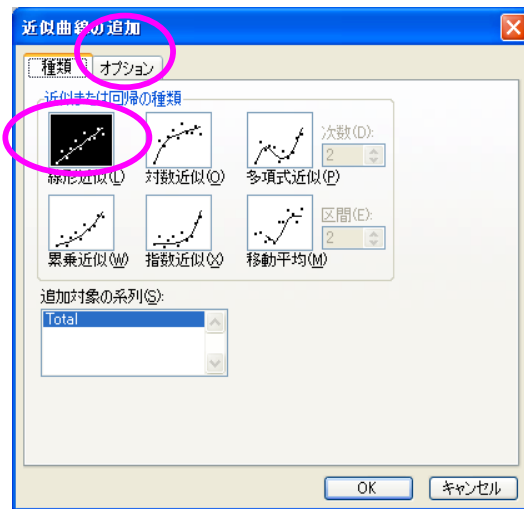
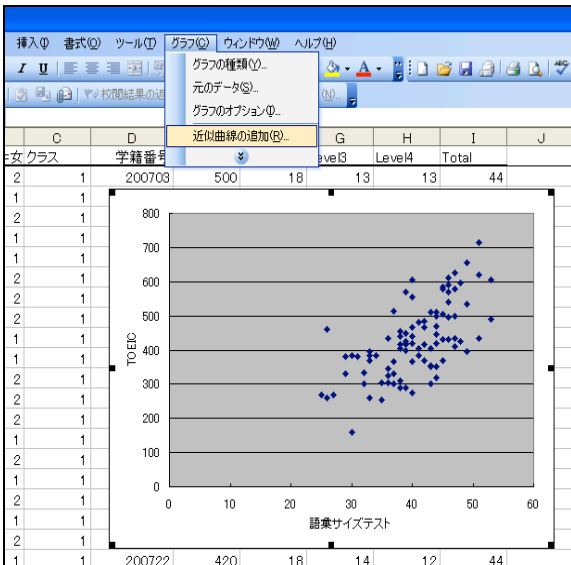
この場合、予測したいテスト (TOEIC) を y 軸にする必要があるため、もう一度、**2.1.2. 散布図を描く** のはじめに戻って、「挿入」⇒「グラフ」を選択。

「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。

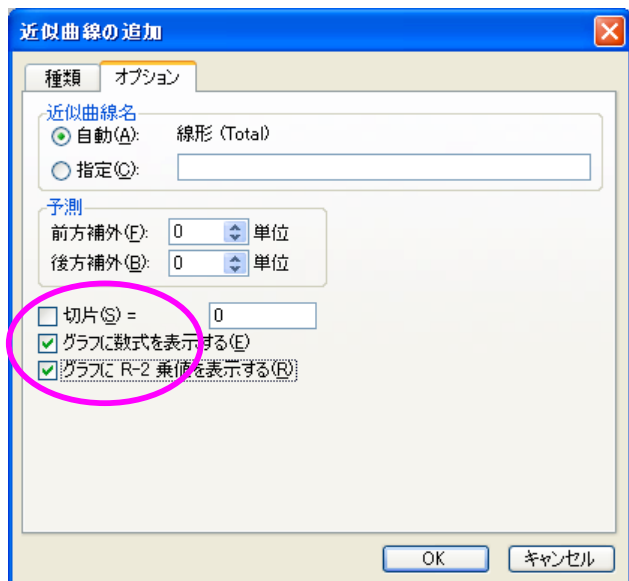


ツールバーの「グラフ」の中から「近似曲線の追加」を選択。「線形近似」が選ばれているのを確認して、「オプション」タブを選択。

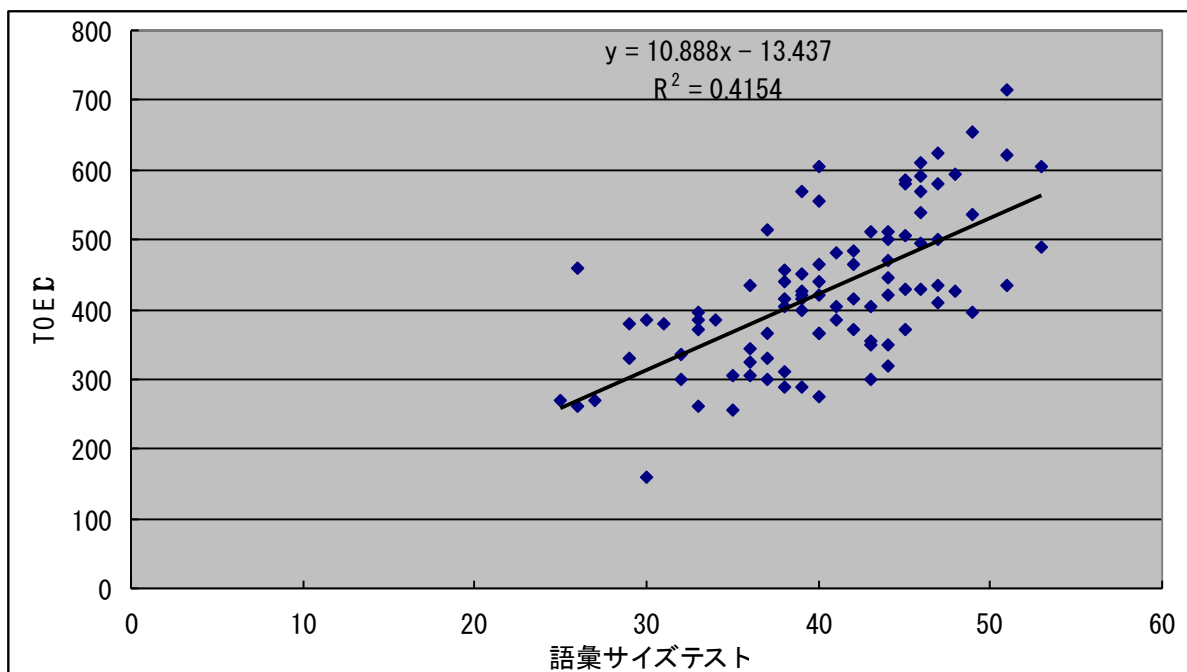


説明使用データ <http://www.mizumot.com/stats/ch9-10.xls>

「オプション」の「グラフに数式を表示する」と「グラフにR²乗値を表示」をチェック



完成。



表の中の数式、 $y = 10.888x - 13.437$ と $R^2 = 0.4154$ は、

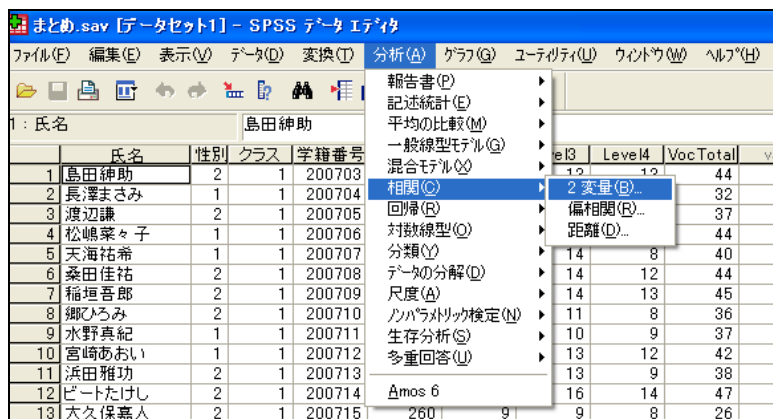
TOEIC の予想点 = (10.888 × 語彙サイズテストの点数) - 13.437
という回帰式が約 **42** パーセントの説明率を持っていることを意味している。

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。

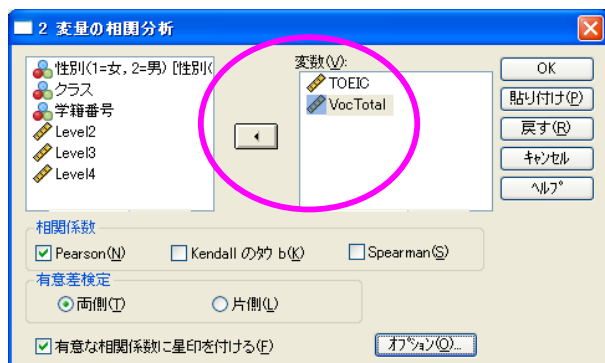
2. SPSS を使って分析する

2.1. 相関係数の計算

「分析」⇒「相関」⇒「2変量」を選ぶ。



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。



結果がアウトプットされる。

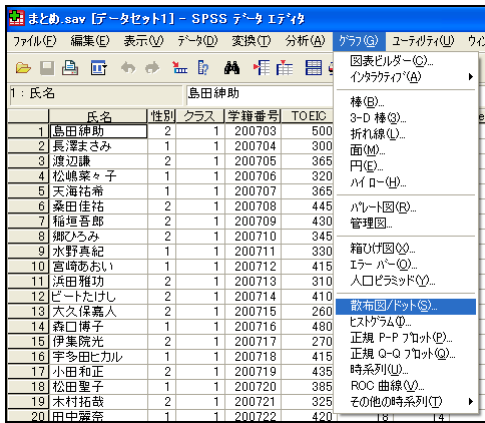
		TOEIC	VocTotal
TOEIC	Pearson の相関係数	1	.645(**)
	有意確率 (両側)		.000
	N	92	92
VocTotal	Pearson の相関係数	.645(**)	1
	有意確率 (両側)	.000	
	N	92	92

** 相関係数は 1% 水準で有意 (両側) です。

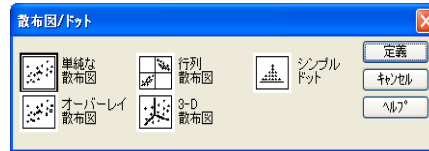
説明使用データ <http://www.mizumot.com/stats/ch9-10.xls>

2.2. 散布図を描く

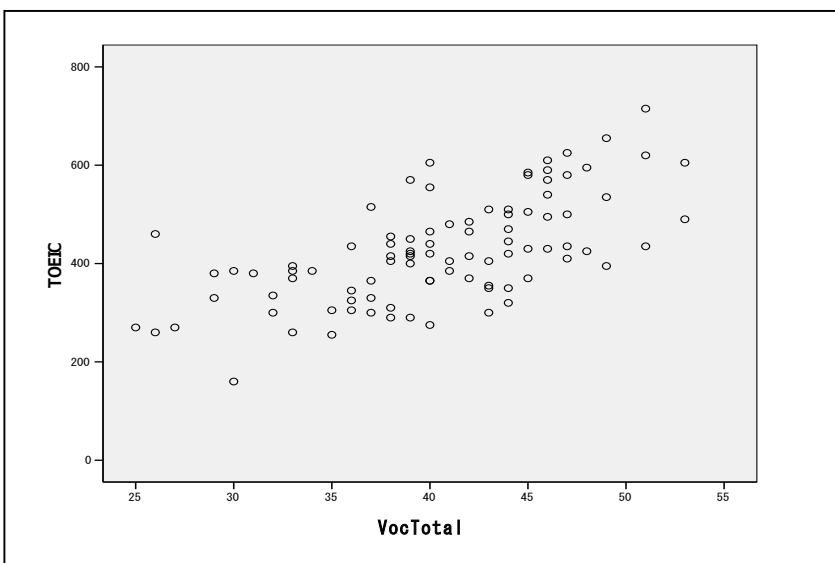
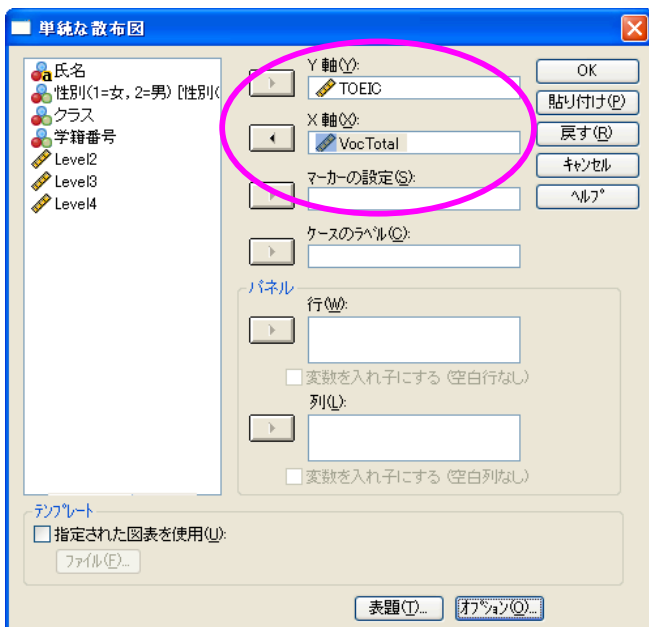
「グラフ」⇒「散布図/ドット」を選び、「単純な散布図」⇒「定義」。



氏名	性別	クラス	学籍番号	TOEIC
島田紳助	2	1	200703	500
長澤まさみ	1	1	200704	300
渡辺謙	2	1	200705	365
松嶋菜々子	1	1	200706	320
天海祐希	1	1	200707	365
森田佳祐	2	1	200708	445
稲垣吾郎	2	1	200709	430
堀ひろみ	2	1	200710	345
水野真紀	1	1	200711	330
宮崎あおい	1	1	200712	415
浜田雅功	2	1	200713	310
ビートたけし	2	1	200714	410
大久保嘉人	2	1	200715	260
森口博子	1	1	200716	480
伊集院光	2	1	200717	270
宇多田ヒカル	1	1	200718	415
小田和正	2	1	200719	435
松田聖子	1	1	200720	385
木村拓哉	2	1	200721	325
田中麗奈	1	1	200722	420



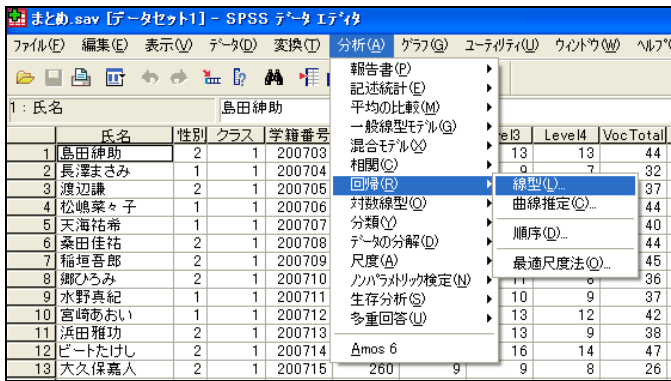
Y 軸, X 軸に変数を選んで移動させて、「OK」をクリック。



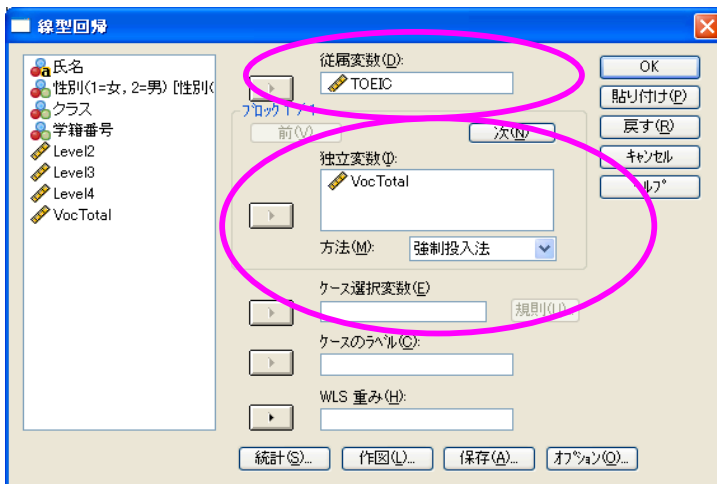
完成。

2.3. 回帰分析をしてみる

「分析」⇒「回帰」⇒「線型」を選ぶ。



「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」。



投入済み変数または除去された変数^a

モデル	投入済み変数	除去された変数	方法
1	VocTotal ^a		投入

a. 必要な変数がすべて投入されました。
b. 従属変数 TOEIC

モデル集計

モデル	R	R ² 乗	調整済み R ² 乗	推定値の標準誤差
1	.645 ^a	.415	.399	62.404

a. 予測値: (定数), VocTotal.

分散分析^b

モデル	回帰	平方和	自由度	平均平方	F 値	有意確率
1	回帰	434863.668	1	434863.668	63.963	.000 ^a
	残差	611882.800	90	6798.698		
	全体	1046746.5	91			

a. 予測値: (定数), VocTotal.
b. 従属変数 TOEIC

係数^a

モデル	非標準化係数	標準化係数	t	有意確率
1	(定数)			
	VocTotal	.645	7.998	.000

a. 従属変数 TOEIC

r (相関係数) = .65
 r^2 (決定係数, 説明率) = .42
 となり, Excel で計算した結果と同じになっている。

$y = 10.888x - 13.437$
 TOEIC 予測点
 $= (10.888 \times \text{語彙サイズテストの点数}) - 13.437$
 という回帰式が求められた。

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。

		TOEIC_total	VocTotal	Motivation
TOEIC_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。

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。

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

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。

モデル	投入済み変数	除去された変数	方法
1	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: TOEIC_total

モデル	R	R ² 乗	調整済み R ² 乗	推定値の標準誤差
1	.656	.430	.417	1.867

a. 予測値: (定数), Motivation, VocTotal.

モデル		平方和	自由度	平均平方	F 値	有意確率
1	回帰	450243.140	2	225121.570	33.589	.000*
	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: TOEIC_total

モデル		非標準化係数		標準化係数	t	有意確率
		B	標準誤差	ベータ		
1	(定数)	-67.671	1.869		-1.027	.307
	VocTotal	10.302	1.406	.610	7.328	.000
	Motivation	19.910	13.144	.126	1.515	.133

a. 従属変数: TOEIC_total

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見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。

$$y = 10.302x_1 + 19.910x_2 - 67.671$$

(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。

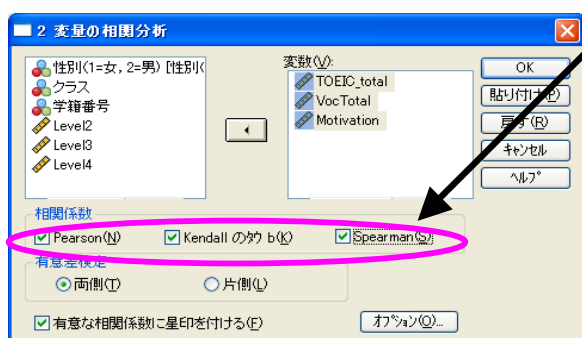
しかし、この部分の有意確率が .05 以上 (.133) なので、「motivation はこの重回帰分析では予測に役立たない」と判断される。

4. 相関係数の種類

- ・ピアソンの積率相関係数 (Pearson'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、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」。

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。