

重回帰分析

B4 現代教育基礎学系 齊藤ユウゴ

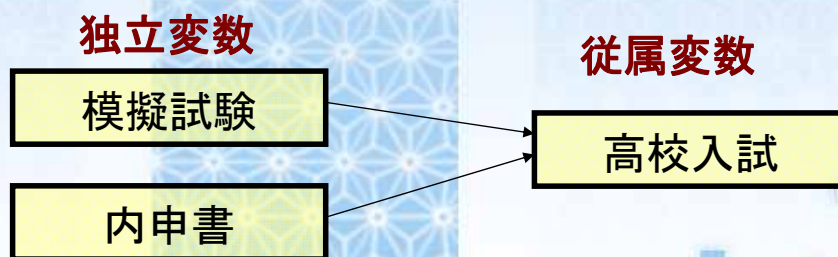
発表の流れ

- 重回帰分析とは何か
- 単回帰分析をやってみましょう
- 単回帰分析の考え方
- 重回帰分析をやってみましょう
- 重回帰分析の考え方
- 重回帰分析を行う上での注意

■重回帰分析とは何か

■**重回帰分析**は、1つの**従属(目的)変数**を、複数の**独立(説明)変数**から予測・説明したいときに用いる統計手法です

例) 中学校の模擬試験と内申書の結果から、高校入試の得点がどの程度説明されるかを検討したい



■重回帰分析とは何か

■ちなみに、

独立変数が1つであれば、『**単回帰分析**』

独立変数が複数であれば、『**重回帰分析**』です

⇒まず、単回帰分析について説明します

例) 中学校の模擬試験の結果から、高校入試の得点がどの程度説明されるかを検討したい。



■ 単回帰分析をやってみましょう

■ 例) 中学校の模擬試験の結果から、
高校入試の得点がどの程度説明されるかを検討したい。

MRsample - SPSS データ エディタ

ファイル(F) 編集(E) 表示(V) データ(D) 変換(T) 分析(A) グラフ(G)

11: 二学期期末テスト

	高校入試	模擬試験	内申書
1	310	322	40
2	412	401	36
3	356	302	45
4	240	301	41
5	423	450	42
6	310	350	38
7	300	285	32
8	215	289	37
9	389	378	40
10	325	310	40
11			
12			

データは、
次のように入力
されています。

※小塩(2004)を参考に
発表者が作成

■ 単回帰分析をやってみましょう

■ 模擬試験と高校入試の間には、どのような関係があるので
しょうか。まず、2つの変数の関係を調べるために、『散布図』を
描いてみます。

MRsample - SPSS データ エディタ

ファイル(F) 編集(E) 表示(V) データ(D) 変換(T) 分析(A) **グラフ(G)** ユーティリティ(U) ウィンドウ(W)

11: 二学期期末テスト

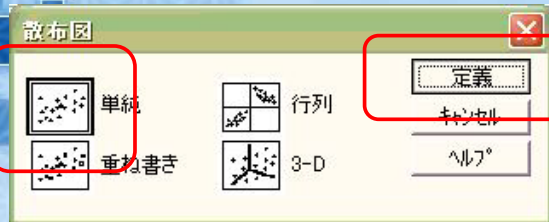
	高校入試	模擬試験	内申書
1	310	322	
2	412	401	
3	356	302	
4	240	301	
5	423	450	
6	310	350	
7	300	285	
8	215	289	
9	389	378	
10	325	310	
11			
12			
13			
14			
15			
16			
17			
18			

棒(E)...
折れ線(L)...
面(M)...
円(C)...
パイロー(H)...
パレート図(R)...
管理図(C)...
箱ひげ図(Q)...
エラーバー(Q)...
散布図(S)...
ヒストグラム(H)...
正規 P-P プロット(P)...
正規 Q-Q プロット(Q)...
時系列(L)...
ROC 曲線(L)...
その他の時系列(T)...

『散布図』の描き方①

「グラフ」メニューの
「散布図」を選択

■ 単回帰分析をやってみましょう

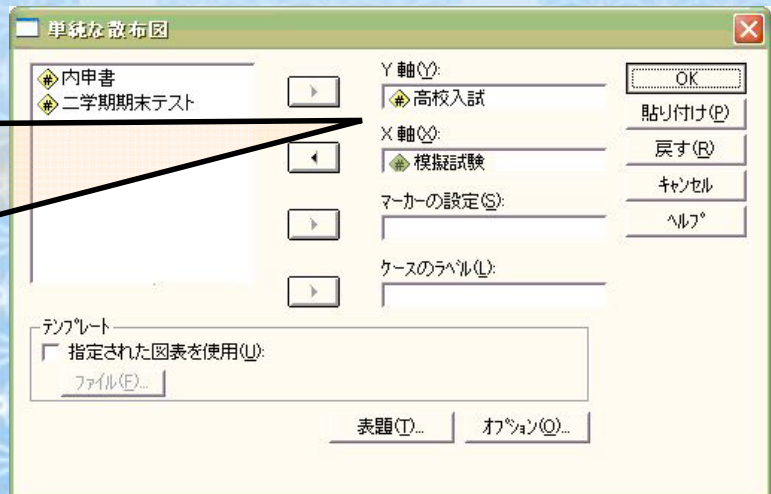


『散布図』の描き方②

「単純」を選択して
「定義」をクリック

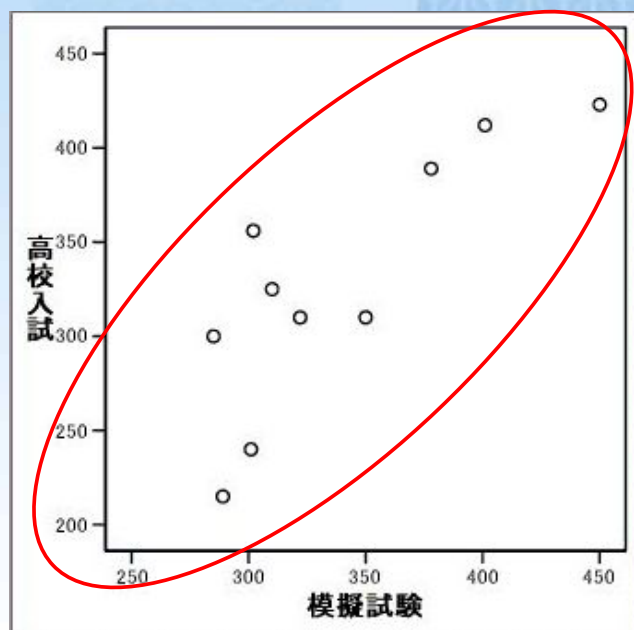
『散布図』の描き方③

Y軸に「高校入試」、
X軸に「模擬試験」を
移動させ、あとは
「OK」を押します



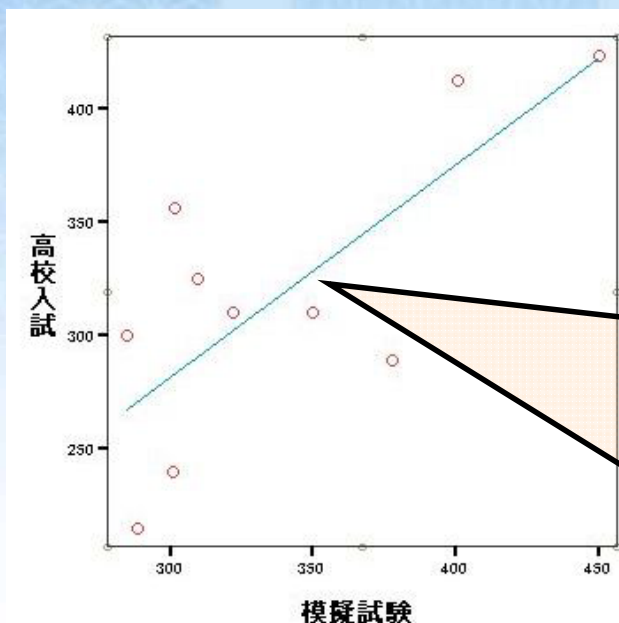
■ 単回帰分析をやってみましょう

■ 次のような、右上がりのデータであることがわかります



■ 単回帰分析をやってみましょう

■ よく見ると、データの点は次の直線のまわりに散らばっているように見えます



このような直線は『**回帰直線**』といい、
 $y = ax + b$
で表すことができます

yは高校入試、
xは模擬試験を表しますので、
高校入試 = a × 模擬試験 + b
のa(傾き)とb(切片)を求めれば
回帰直線の式を求めることができます

■ 単回帰分析をやってみましょう

■ 実際に、単回帰式のaとbの値を、SPSSを使って単回帰分析をやって、求めてみましょう

MRsample - SPSS データエディタ

ファイル(F) 編集(E) 表示(V) データ(D) 変換(C) 分析(A) グラフ(G) ユーティリティ(U) ウィンドウ(W) ヘルプ(H)

11: 二学期期末テスト

	高校入試	模擬試験
1	310	322
2	412	401
3	356	302
4	240	301
5	423	450
6	310	350
7	300	285
8	215	289
9	389	378
10	325	310
11		
12		
13		
14		

分析(A) メニュー:

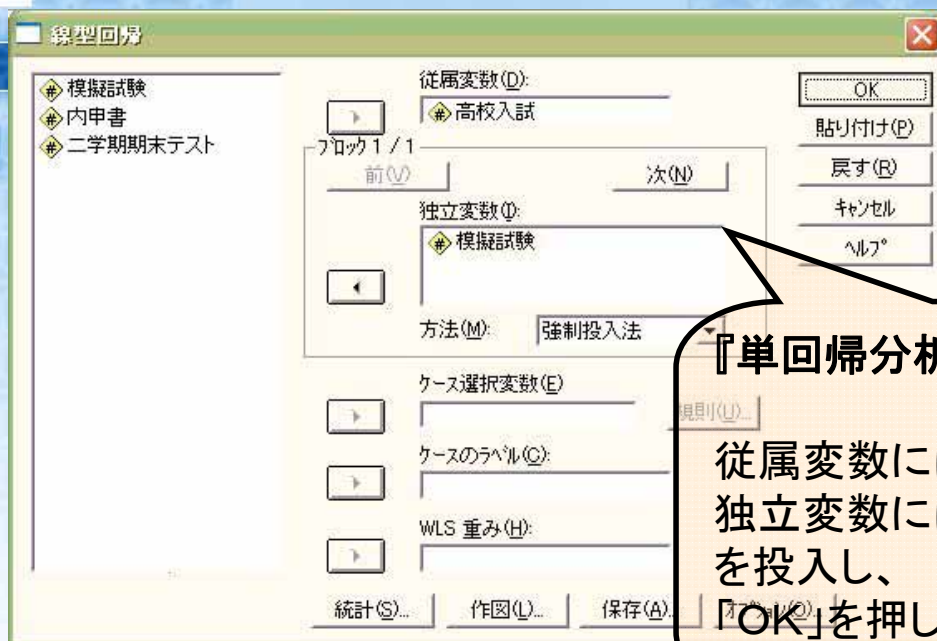
- 報告書(P)
- 記述統計(E)
- 平均の比較(M)
- 一般線型モデル(Q)
- 混合モデル(O)
- 相関(C)
- 回帰(R)
- 対数線型(O)
- 分類(Y)
- データの分解(D)
- 尺度(A)
- ノンパラメトリック検定(N)
- 生存分析(S)
- 多重回答(U)

線型(L)...

『単回帰分析』手順①

「分析」メニューの
「回帰」から
「線型」を選択

■ 単回帰分析をやってみましょう



線型回帰

従属変数(D): 高校入試

独立変数(I): 模擬試験

方法(M): 強制投入法

統計(S)... 作図(L)... 保存(A)...

OK 貼り付け(P) 戻す(R) キャンセル ヘルプ

『単回帰分析』手順②

従属変数には「高校入試」、
独立変数には「模擬試験」
を投入し、
「OK」を押します

■ 単回帰分析をやってみましょう

係数^a

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)	-18.675	87.303		-.214	.836
	模擬試験	1.023	.255	.818	4.018	.004

a. 従属変数: 高校入試

読み取るのは、
この2つです！
は傾きa
は切片b
をあらわします

この結果から、単回帰分析の単回帰式は

$$y = 1.023x - 18.675$$

高校入試 $1.023 \times$

模擬試験

であることがわかります

■ 単回帰分析をやってみましょう

★今度は、中学校の内申書から、
高校入試の得点がどの程度説明されるかを検討してください



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) 176.777	251.814		.702	.503
	内申書 3.612	6.416	.195	.563	.589

a. 従属変数: 高校入試

この結果から、単回帰分析の単回帰式は

$$y = 3.621x - 176.777$$

高校入試 $3.621 \times$ 内申書 -176.777

であることがわかります

■ 単回帰分析の考え方

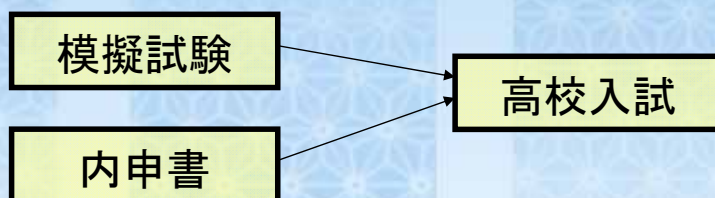
■ 単回帰式はどのようにして求められるのかを(時間があれば)解説します

■ キーワード

- ① 最小2乗法
- ② 分散と共分散
- ③ 2次関数の最小値

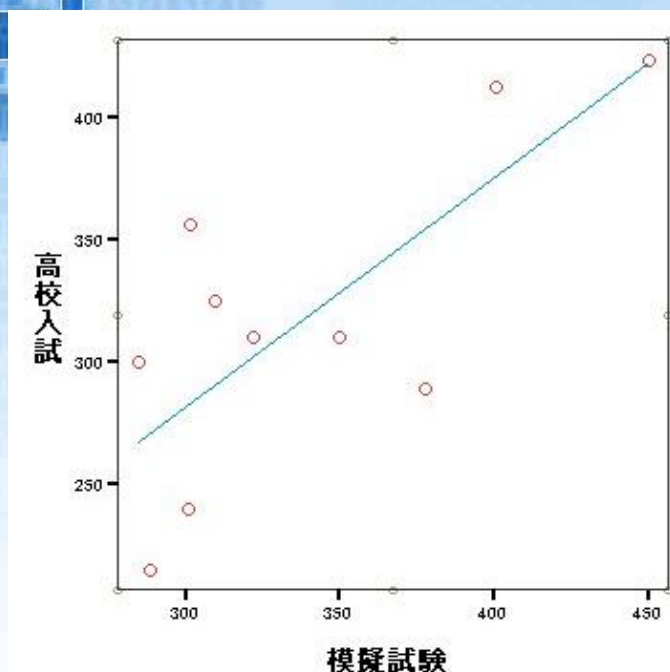
■重回帰分析をやってみましょう

■例) 中学校の模擬試験と内申書の結果から、高校入試の得点がどの程度説明されるかを検討したい



■単回帰分析のときは、
高校入試 y と模擬試験 x_1 の間に $y=a_1x_1+b_1$
高校入試 y と内申書 x_2 の間に $y=a_2x_2+b_2$
という1次式の関係があると考えましたが、今度は
2つの式をまとめて、 $y=a_1x_1+a_2x_2+a_0$
という1次式を求めることにします

■重回帰分析をやってみましょう



■ただ、左の回帰直線を見てもわかるように、各データの点はきちんと直線上にあるわけではありません

そこで、はじめに与えられているデータを

実測値 y (スモールワイ)、
1次式で(計算して)与えられる値を

予測値 Y (ラージワイ)
で表すことにします

■重回帰分析をやってみましょう

MRsample - SPSS データエディタ

ファイル(F) 編集(E) 表示(V) データ(D) 変換(T) 分析(A) グラフ(G)

11: 二学期期末テスト

	高校入試	模擬試験	内申書
1	310	322	40
2	412	401	36
3	356	302	45
4	240	301	41
5	423	350	42
6	310	285	38
7	300	289	32
8	215	378	37
9	389	310	40
10	325		40
11			
12			
13			

これは
実測値です

■つまり、
実測値…y

予測値… $Y = a_1X_1 + a_2X_2 + a_0$

となります

さらに、予測値と実測値のズレを
『残差』と呼び、

実測値－予測値… $y - Y$

で表すことになります

■で、求めなければならないものは、予測値Yを与える
重回帰式 $Y = a_1X_1 + a_2X_2 + a_0$ です。独立変数の係数
 a_1 、 a_2 のことを偏回帰係数、 a_0 を定数項といいます。

■重回帰分析をやってみましょう

MRsample - SPSS データエディタ

ファイル(F) 編集(E) 表示(V) データ(D) 変換(T) 分析(A) グラフ(G) ユーティリティ(U) ウィンドウ(W) ヘルプ(H)

1: 高校入試 310

	高校入試	模擬試験
1	310	322
2	412	401
3	356	302
4	240	301
5	423	450
6	310	350
7	300	285
8	215	289
9	289	378
10	325	310
11		
12		
13		

分析(A) メニュー:

- 報告書(P)
- 記述統計(E)
- 平均の比較(M)
- 一般線型モデル(G)
- 混合モデル(O)
- 相関(C)
- 回帰(R)
- 対数線型(O)
- 分類(Y)
- データの分解(D)
- 尺度(A)
- ノンパラメトリック検定(N)
- 生存分析(S)
- 多重回答(U)

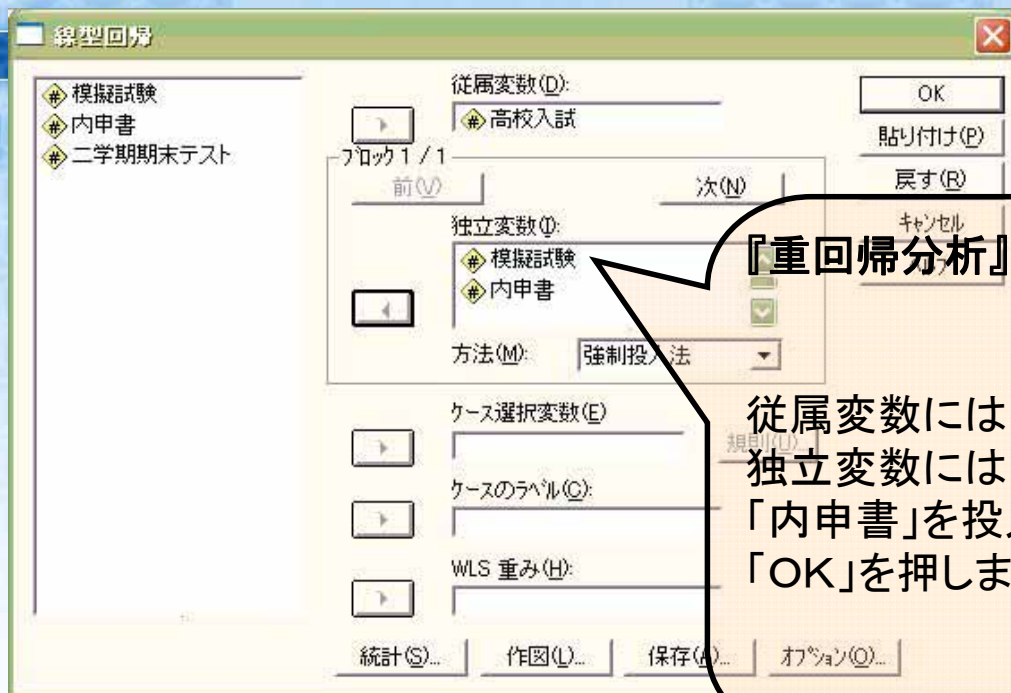
回帰(R) のサブメニュー:

- 線型(L)...
- 曲線推定(C)...
- 二項ロジスティック(G)...
- 多項ロジスティック(M)...
- 順序(O)...
- プロビット(P)...
- 非線型(N)...
- 重み付き最小二乗(WLS)...
- 2段階最小二乗(2SLS)...

『単回帰分析』手順①

「分析」メニューの
「回帰」から
「線型」を選択

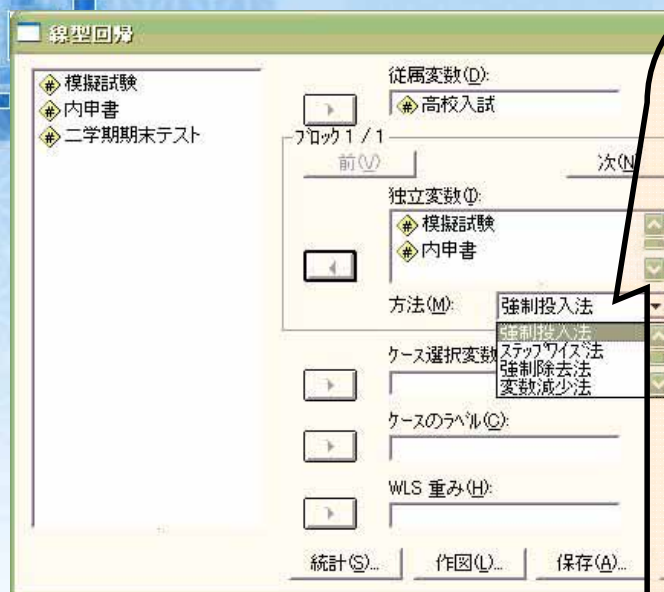
■重回帰分析をやってみましょう



『重回帰分析』手順②

従属変数には「高校入試」、
独立変数には「模擬試験」と
「内申書」を投入し、
「OK」を押します

■重回帰分析をやってみましょう



『重回帰分析』手順補

『ステップワイズ法』

既存のモデルをもとに、次に新しい変数を入れるか、あるいはモデルに既に入っている変数を落とすかを逐次的に行う。

『変数増加法』

予測の精度を高める変数の順に追加。

『変数減少法』

予測の精度に高めない変数を順に削除。

■重回帰分析をやってみましょう

係数 ^a					
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)	-78.357	174.536	-.449	.667
	模擬試験	1.004	.273	.802	.008
	内申書	1.693	4.204	.088	.699

a. 従属変数: 高校入試

読み取るのは、
この3つです！

□ は
偏回帰係数、
□ は定数項
を表します

この結果から、重回帰分析の重回帰式は

$$Y = 1.004 X_1 + 1.693 X_2 - 78.357$$

$$Y = 1.004 \times [\text{模擬試験}] + 1.693 \times [\text{内申書}] - 78.357$$

であることがわかります

■重回帰分析の考え方

■偏回帰係数は何を意味するのか

■キーワード

①重相関係数と決定係数

②標準(偏)回帰係数

■重回帰分析の考え方

■①重相関係数と決定係数を調べる

さきほど求めた重回帰式

$$Y = 1.004 \times [\text{模擬試験}] + 1.693 \times [\text{内申書}] - 78.357$$

に、模擬試験の点数、内申書の点数を代入すると、
高校入試の点数Yを予測することができます。

しかし、予測が外れることもあります。

その原因として、**重回帰式の当てはまりのわるさ**があります。

■重回帰分析の考え方

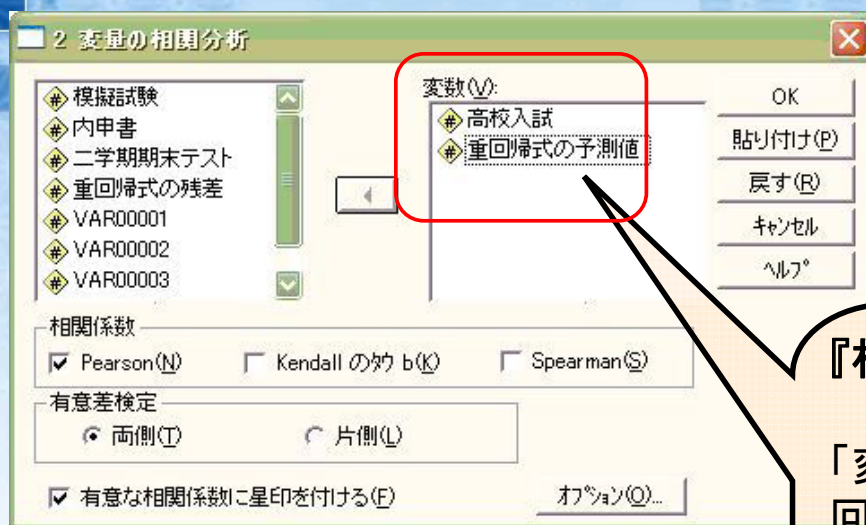
The screenshot shows the SPSS 'Analyze' menu with the following path highlighted by red boxes: **分析(A)** → **相関(C)** → **2変量(B)...**. The background shows a data table with columns for '高校入試' (High School Entrance Exam) and '模擬試験' (Mock Exam).

	高校入試	模擬試験
1	310	322
2	412	401
3	356	302
4	240	301
5	423	450
6	310	350
7	300	285
8	215	289
9	389	378
10	325	310
11		
12		
13		

『相関係数』の求め方①

「分析」メニューの
「相関」→「2変量」
を選択

■重回帰分析の考え方



『相関係数』の求め方②

「変数」に「高校入試」と「重回帰式の予測値」を投入して「OK」を押します

■重回帰分析の考え方

相関係数			
		高校入試	重回帰式の予測値
高校入試	Pearson の相関係数	1	.822**
	有意確率 (両側)		.004
	N	10	
重回帰式の予測値	Pearson の相関係数	.822**	
	有意確率 (両側)	.004	
	N	10	

** 相関係数は 1% 水準で有意 (両側) です。

『相関係数』の求め方③

「高校入試」と「重回帰式の予測値」の相関係数は 0.822 で 1% 水準で有意です

■重回帰分析では、この相関係数のことを『**重相関係数R**』と呼んでいます。

■重回帰分析の考え方

■重相関係数Rは

$$0 \leq R \leq 1$$

が成り立ちます。今回の重相関係数は0.822なので、1に近く、求めた重回帰式はデータにととてもよく当てはまっていると考えられます。

■重相関係数Rの2乗を『**決定係数(寄与率)R²**』といいます。

$$0 \leq R^2 \leq 1$$

決定係数は、『実測値』『予測値』『残差』の分散で求めることもできます。
が成り立ち、やはり**決定係数R²**も1に近いほうが**重回帰式**がデータによく当てはまっているといえます。

■重回帰分析の考え方

■重相関係数Rも、決定係数R²も、SPSSで重回帰分析をすると、自動的に出力されます！

モデル集計			
モデル	R	R ² 乗	調整済みR ² 乗
1	.822 ^a	.676	.584

a. 予測値: (定数)、内申書 模擬試験。

重相関係数

決定係数

決定係数R²は回帰方程式のあてはまりのよさを示しますが、独立変数の数を増やすと決定係数は単純に増加し、見かけ上、当てはまりがよくなってしまいます。この欠点を補うために**自由度調整済み決定係数R²**を用います。

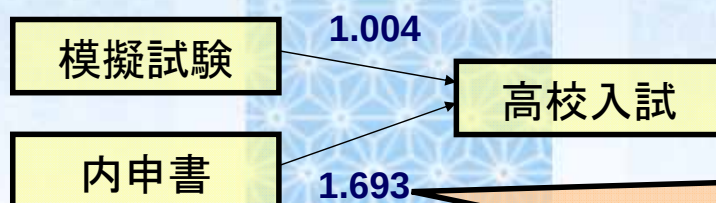
■重回帰分析の考え方

■②標準(偏)回帰係数について

求めた重回帰式

$$Y = 1.004 \times [\text{模擬試験}] + 1.693 \times [\text{内申書}] - 78.357$$

ですが、模擬試験と内申書、どちらが高校入試の点数に、より大きな影響を与えているのでしょうか？



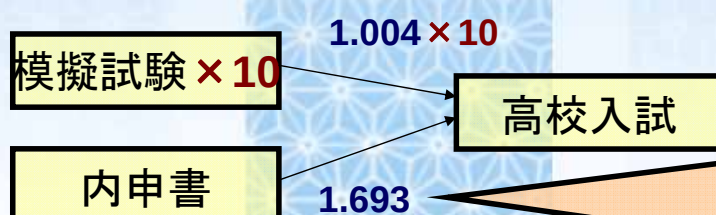
内申書のほうが偏回帰係数が大きいので内申書のほうが影響を与えている……**わけではありません！！**

■重回帰分析の考え方

■なぜならば、模擬試験と内申書は

『ものさし』が違うからです！

たとえば模擬試験が5000点満点で、全員の模擬試験の点数が10倍になったら、模擬試験の偏回帰係数は10倍の10.04になります。



つまり、『どちらのほうが影響を与えているか』をしらべるには、偏回帰係数ではだめです。

■重回帰分析の考え方

■『ものさし』によらないように比べるためには⇒**データの標準化**を行います！

すべてのデータを標準化し、同じように重回帰分析を行えば、『ものさし』によらない**偏回帰係数**、

『**標準(偏)回帰係数**』を出すことができます

…が、実はSPSSに自動的に出力されてます。

係数 ^a					
モデル		非標準化係数		標準化係数	有意確率
		B	標準誤差	ベータ	
1	(定数)	-78.357	174.536		.667
	模擬試験	1.004	.273	.802	.008
	内申書	1.693	4.204	.088	.699

a. 従属変数: 高校入試

これがデータの『ものさし』によらない係数、**標準(偏)回帰係数**です

■重回帰分析の考え方

■「模擬試験」と「内申書」の標準偏回帰係数を比べると、模擬試験の割合のほうが大きいです。

係数 ^a				
モデル		非標準化係数		標準化係数
		B	標準誤差	ベータ
1	(定数)	-78.357	174.536	
	模擬試験	1.004	.273	.802
	内申書	1.693	4.204	.088

a. 従属変数: 高校入試

このことは、「高校入試の得点に及ぼす影響の強さは、内申書よりも模擬試験の割合のほうが大きい」ことを示しています。

■重回帰分析の考え方

■SPSSで重回帰分析を行った際、出力をどのように見たらいいかをまとめます。

①まず、モデルの適合度を見ます(注目するのは**自由度調整済み決定係数R²値**)

モデル集計 ^a				
モデル	R	R ² 乗	調整済みR ² 乗	推定値の標準誤差
1	.822 ^a	.676	.584	44.665

a. 予測値: (定数), 内申書, 模擬試験。

今回、模擬試験と内申書で、高校入試の得点の58.4%を説明できることを示しています。

■重回帰分析の考え方

■SPSSで重回帰分析を行った際、出力をどのように見たらいいかをまとめます。

②次に、分散分析のF値の有意確率をみます(モデルで用いた独立変数で従属変数を説明することが妥当かどうかを検定してくれています)

分散分析 ^b						
モデル		平方和	自由度	平均平方	F 値	有意確率
1	回帰	28762.476	2	14381.238	7.307	.019 ^a

今回、F値はモデルが有意水準5%以下で妥当であることを示しています

今回、F値はモデルが有意水準5%以下で妥当であることを示しています

■重回帰分析の考え方

■SPSSで重回帰分析を行った際、出力をどのように見たいかをまとめます。

③最後に、「非標準化係数(偏回帰係数)」、「標準化係数(標準偏回帰係数)」、「有意確率」を確認します。

今回、模擬試験は1%水準で高校入試得点に統計的に有意な影響を与えていることを示しています。

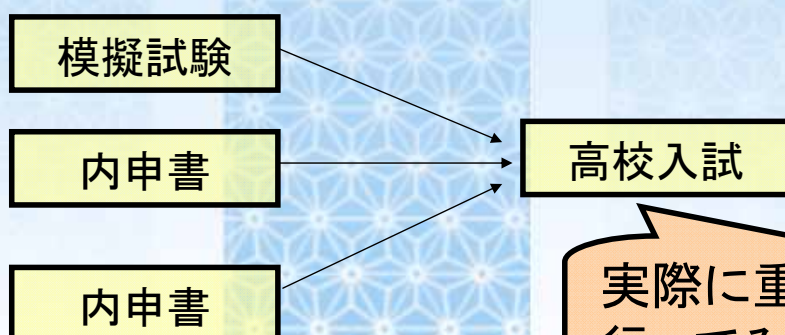
係数 ^a					
モデル		非標準化係数		標準化係数	t
		B	標準誤差	ベータ	
1	(定数)	-78.357	174.536		-.449
	模擬試験	1.004	.273	.802	3.672
	内申書	1.693	4.204	.088	.403

a. 従属変数: 高校入試

■重回帰分析を行う上での注意

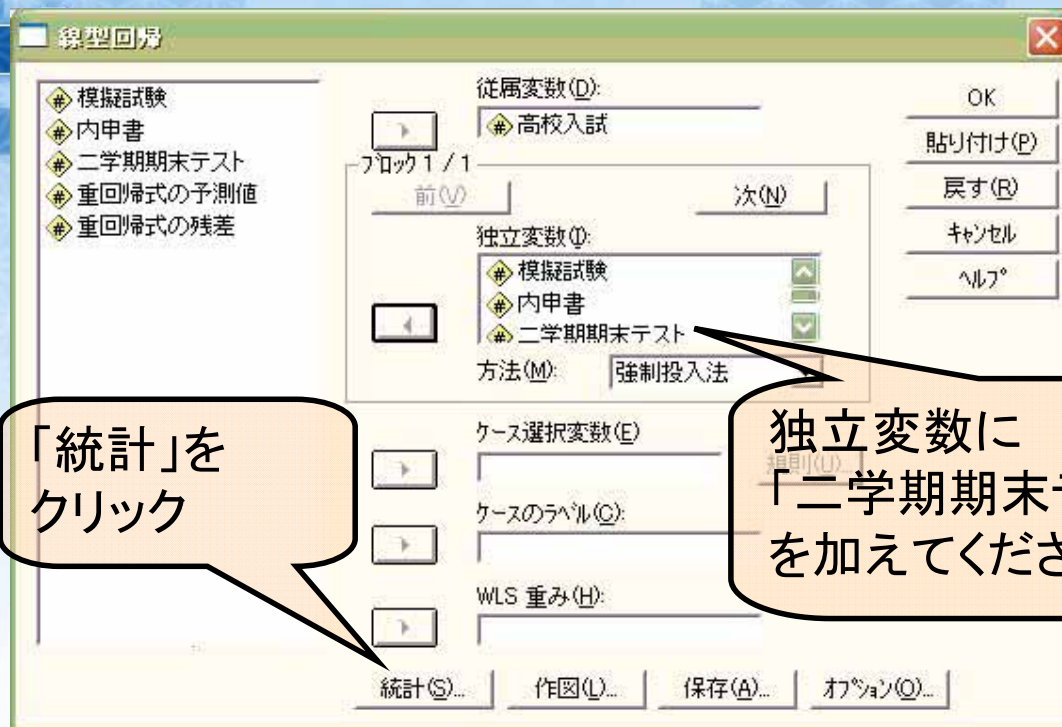
■多重共線性(マルチコ)について

例) 中学校の模擬試験と内申書と二学期期末テストの結果から、高校入試の得点がどの程度説明されるかを検討したい

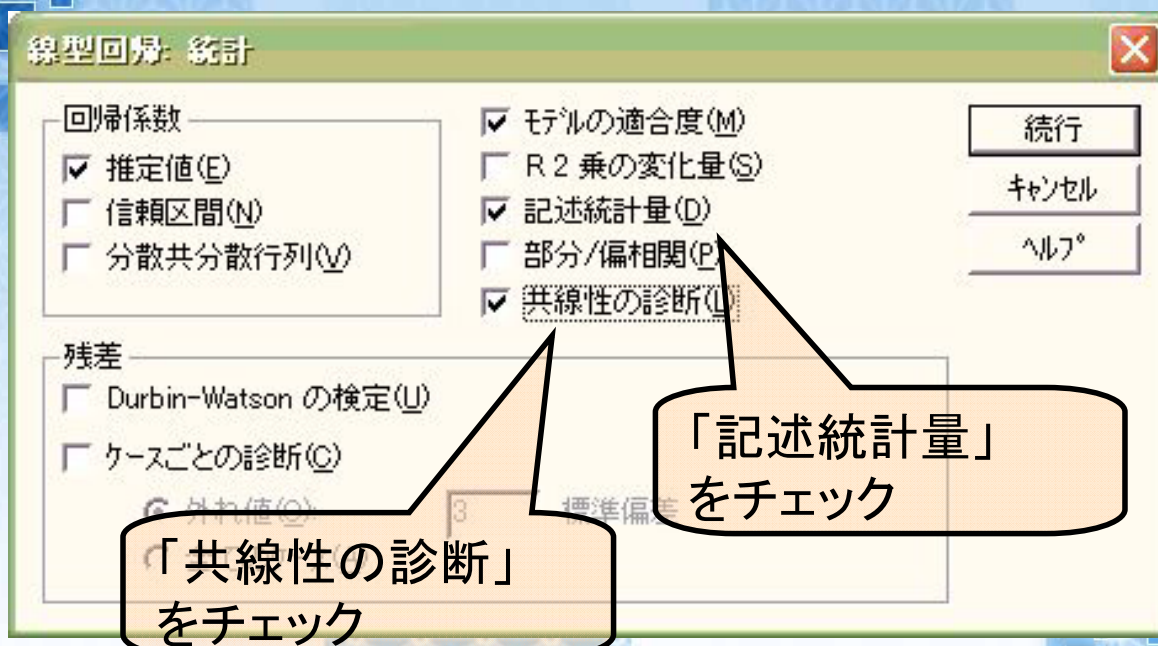


実際に重回帰分析を行ってください

■重回帰分析を行う上での注意



■重回帰分析を行う上での注意



■重回帰分析を行う上での注意

係数 ^a								
モデル		非標準化係数		標準化係数	t	有意確率	共線性の統計量	
		B	標準誤差	ベータ			許容度	VIF
1	(定数)	2471.229	1715.294		1.441	.200		
	模擬試験	1.371	.352	1.095	3.894	.008	.497	2.011
	内申書	58.145	38.010	3.022	1.530	.177	.010	99.149
	二学期期末テスト	-11.739	7.863	-3.014	-1.493	.186	.010	103.527

a. 従属変数: 高校入試

「二学期期末テスト」の偏
回帰係数の符号が**マイナ
ス**に！

■予想とだいぶちがう！

普通に考えれば、二学期期末
テストの得点が高かった生徒
は、高校試験の得点も高くな
るのでは？

■重回帰分析を行う上での注意

■多重共線性が発生しているかも！

相関係数を確認すると、
ともに独立変数の
「内申書」と「二学期期末テ
スト」は、0.991という非常に
強い相関が！！

⇒独立変数間に強い相
関があるときは
要注意！

相関係数

高校入試	模擬試験	内申書	二学期期 末テスト
1.000	.818	.228	.273
.818	1.000	.175	.268
.228	.175	1.000	.991
.273	.268	.991	1.000
.002	.002	.263	.223
.002	.002	.314	.227
.263	.314	.000	.000
.223	.227	.000	.000
10	10	10	10
10	10	10	10
10	10	10	10
10	10	10	10

■重回帰分析を行う上での注意

■多重共線性が発生しているかも！

このように独立変数間に強い相関がある(線形関係が成り立つ)とき、回帰係数の符号が予想と反したり、あきらかにおかしい値になったりすることがある

⇒多重共線性(multi-collinearity)という

SPSSでは多重共線性が発生してるかどうかを判断する指標として、
VIF (Variance Inflation Factor)
を算出することができる

■重回帰分析を行う上での注意

■指標といっても絶対的なものではないですが

係数 ^a								
モデル		非標準化係数		標準化係数	t	有意確率	共線性の統計量	
		B	標準誤差	ベータ			許容度	VIF
1	(定数)	2471.229	1715.294		1.441	.200		
	模擬試験	1.371	.352	1.095	3.894	.008	.497	2.011
	内申書	58.145	38.010	3.022	1.530	.177	.010	99.149
	二学期期末テスト	11.755	7.603	0.614	1.535	.186	.010	103.527

VIFが高い値(10以上だったり2以上だったりと文献によってさまざま)を示している独立変数は、他の独立変数と相関が強い可能性があります

■重回帰分析を行う上での注意

■多重共線性が発生した場合の対処

①少なくとも1つの独立変数(たとえば重要度の低い変数)を削除する

(今回の場合、「内申書」か「二学期期末テスト」どちらかを削除)

②相関の高い複数の独立変数を主成分分析(あるいは因子分析)により合成する

■参考文献

□小塩真司(2004)『SPSSとAmosによる心理・調査データ解析—因子分析・共分散構造分析まで—』東京図書株式会社

□石村貞夫・石村光資郎(1999)『SPSSでやさしく学ぶ多変量解析』東京図書株式会社

□丹慶勝市(2007)『多変量解析』株式会社ナツメ社