

重回帰分析 (Multiple Regression Analysis)

$$y = a + bx$$

独立変数が一つ＝単回帰分析

$$y = a + b_1x_1 + b_2x_2 + \cdots + b_{n-1}x_{n-1} + b_nx_n$$

独立変数が複数＝重回帰分析

。

○

○

複雑な式がでてきたと思われるかもしれないが、
 x_1 、 x_2 、 x_3 、...、 x_{n-1} 、 x_n といった複数の独立変数が、
従属変数 y の値を決定するという関係を表したもの

x_1 が 1 単位変化すれば y が b_1 変化すること、

x_2 が 1 単位変化すれば y は b_2 変化すること、

：

x_n が 1 単位変化すれば y は b_n 変化することを意味する

(それぞれその他の独立変数の値が不変として)

重回帰分析の実際

仮説:「成績は講義への出席率、入試成績によって決定される」

独立変数＝出席率、入試成績



従属変数＝成績

区分	成績 Y (点)	出席率 X_1 (%)	入試成績 ₂ (点)
A	65	75	70
B	70	65	70
C	90	90	85
D	70	65	80
E	75	75	80
F	85	90	85
G	90	90	90
H	85	75	80
I	75	80	75
J	60	55	70
K	55	65	40
L	70	80	75
M	75	70	70
N	90	90	95
O	80	85	80

*無題1 [データセット0] - IBM SPSS Statistics 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) ダイレクト マーケティング(M) グラフ(G) ユーティリティ(U) ウィンドウ(W) ヘルプ(H)

表示: 3 個 (3 変数中)

	成績	出席率	入試成績	var	var	var	var	var	var	var	var	var	var	var	var	var
1	65	75	70													
2	70	65	70													
3	90	90	85													
4	70	65	80													
5	75	75	80													
6	85	90	85													
7	90	90	90													
8	85	75	80													
9	75	80	75													
10	60	55	70													
11	55	65	40													
12	70	80	75													
13	75	70	70													
14	90	90	95													
15	80	85	80													
16																
17																
18																
19																
20																
21																
22																
23																
24																

データ ビュー(D) 変数 ビュー(V)

線型回帰

統計量(S)...
作図(T)...
保存(S)...
オプション(O)...

従属変数(D): 成績

ブロック 1 / 1

前(V) 次(N)

独立変数(I): 出席率
入試成績

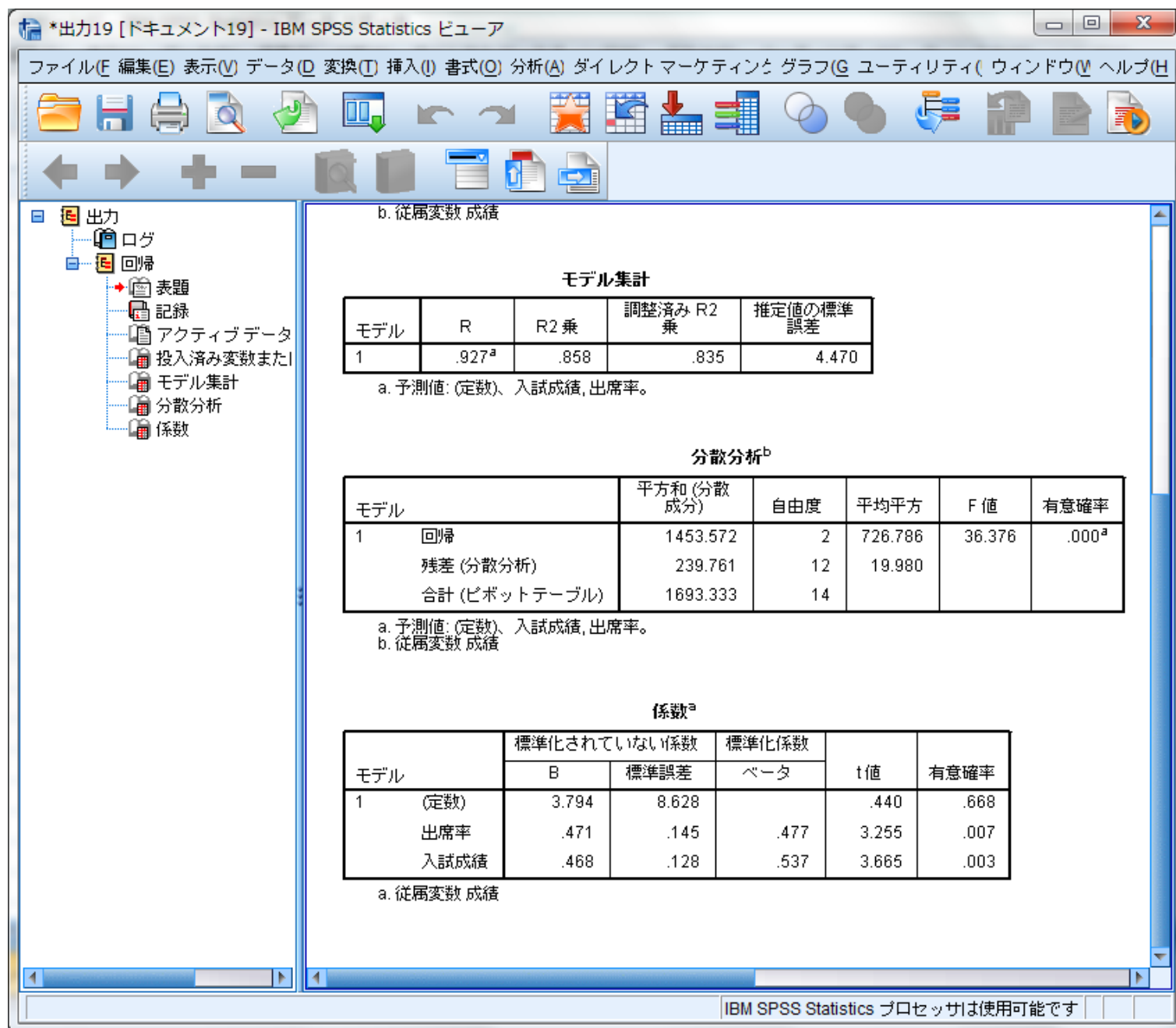
方法(M): 強制投入法

ケース選択変数(E) 規則(U)...

ケースのラベル(C):

WLS 重み(H):

OK 貼り付け(P) 戻す(R) キャンセル ヘルプ



計算結果をもとに重回帰式を整理すると、

$$Y = 3.794 + 0.471X_1 + 0.468X_2$$

しかし、問題は残る。たとえば、

これってほとんどゼロじゃないの？
この係数に意味があるの？

たとえば X_1 の係数が0.01となった時に、講義への出席率が1%上がれば、
成績は0.01点高くなると言ってよいのだろうか？
(係数に意味があるのだろうか)

t値⇒個々の独立変数の係数（偏回帰係数）を検証するもの

帰無仮説「偏回帰係数＝0」を検定

$$Y = 3.794 + 0.471X_1 + 0.468X_2$$

(0.44) (3.26) (3.67)

F値⇒回帰方程式全体を検証するためのもの

→複数の独立変数を全体として一括して検証

帰無仮説「Y=0」を検定

ダミー変数(dummy variable)

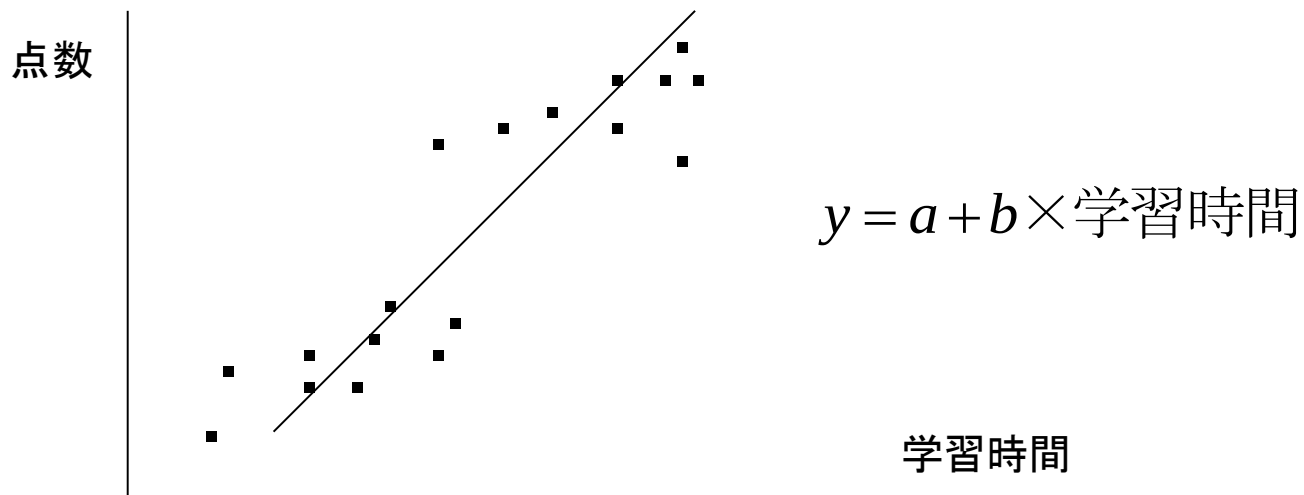
ダミー変数⇒質的変数(名義尺度、順序尺度)を回帰分析に用いるための手法

仮説:

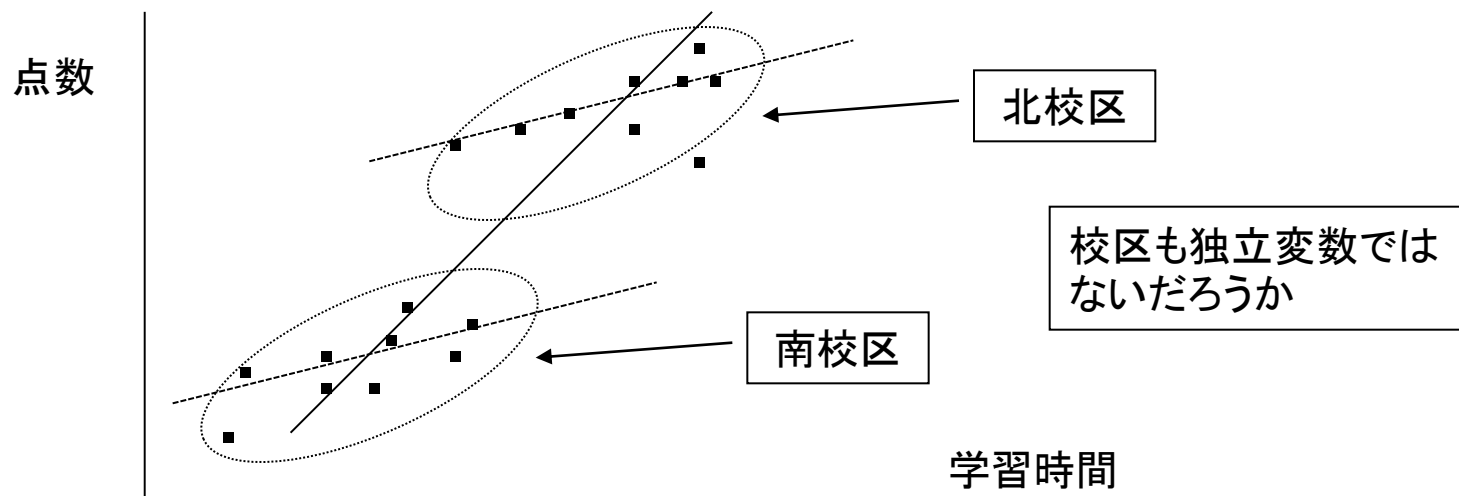
独立変数＝学習時間



従属変数＝点数



しかし、決定係数が高くない



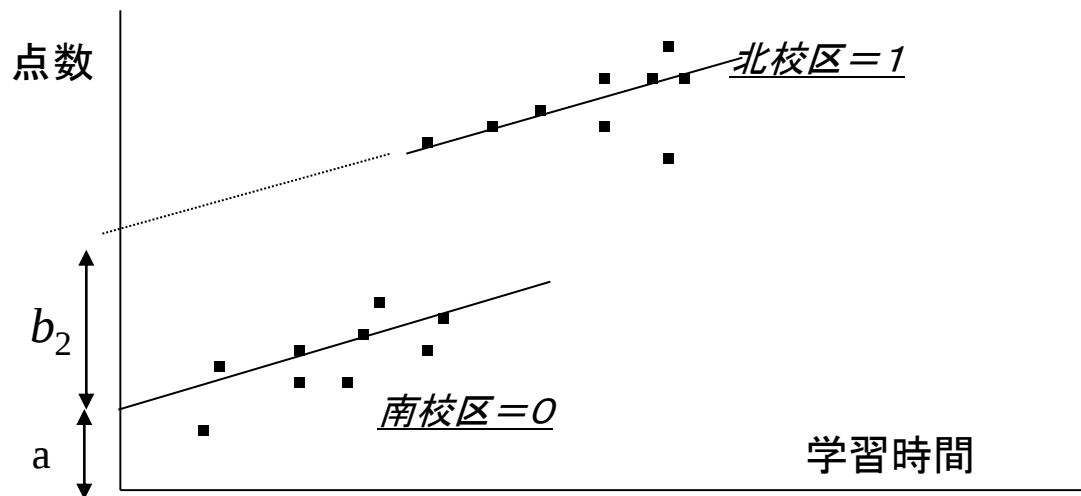
仮説:

独立変数 = 学習時間、校区



従属変数 = 点数

校区をダミー変数(北校区=1、南校区=0)として回帰分析を行う



$$y = a + b_1 \times \text{学習時間} + b_2 \times \text{校区}$$

ダミー変数の入力

質的変数の例： 性別（男、女）

人種（white, African America, Asian, Hispanic...）

都市（都市部、農村部）

政党（conservative, liberal）

学歴（大卒以上、高卒以下）などなど

入力＝ゼロ・サム方式で入力

例： 男＝1、女＝0 （もちろん男＝0、女＝1でもよい）

白人＝1、非白人＝0 （もちろん逆も可、以下同）

都市部＝1、農村部＝0

保守＝1、革新＝0

大卒以上＝1、高卒以下＝0

① SPSSによる重回帰分析

仮説:

独立変数＝出席率、入試成績、教科書購入、性別



従属変数＝成績

区分	成績 Y (点)	出席率 X ₁ (%)	入試成績 X ₂ (点)	教科書購入 の有無 X ₃	性別 X ₄
A	65	75	70	無	男
B	70	65	70	有	男
C	90	90	85	有	女
D	70	65	80	有	男
E	75	75	80	有	男
F	85	90	85	有	女
G	90	90	90	有	女
H	85	75	80	有	女
I	75	80	75	有	女
J	60	55	70	無	男
K	55	65	40	無	男
L	70	80	75	無	男
M	75	70	70	有	女
N	90	90	95	有	女
O	80	85	80	有	男

*無題1 [データセット0] - IBM SPSS Statistics データ エディタ

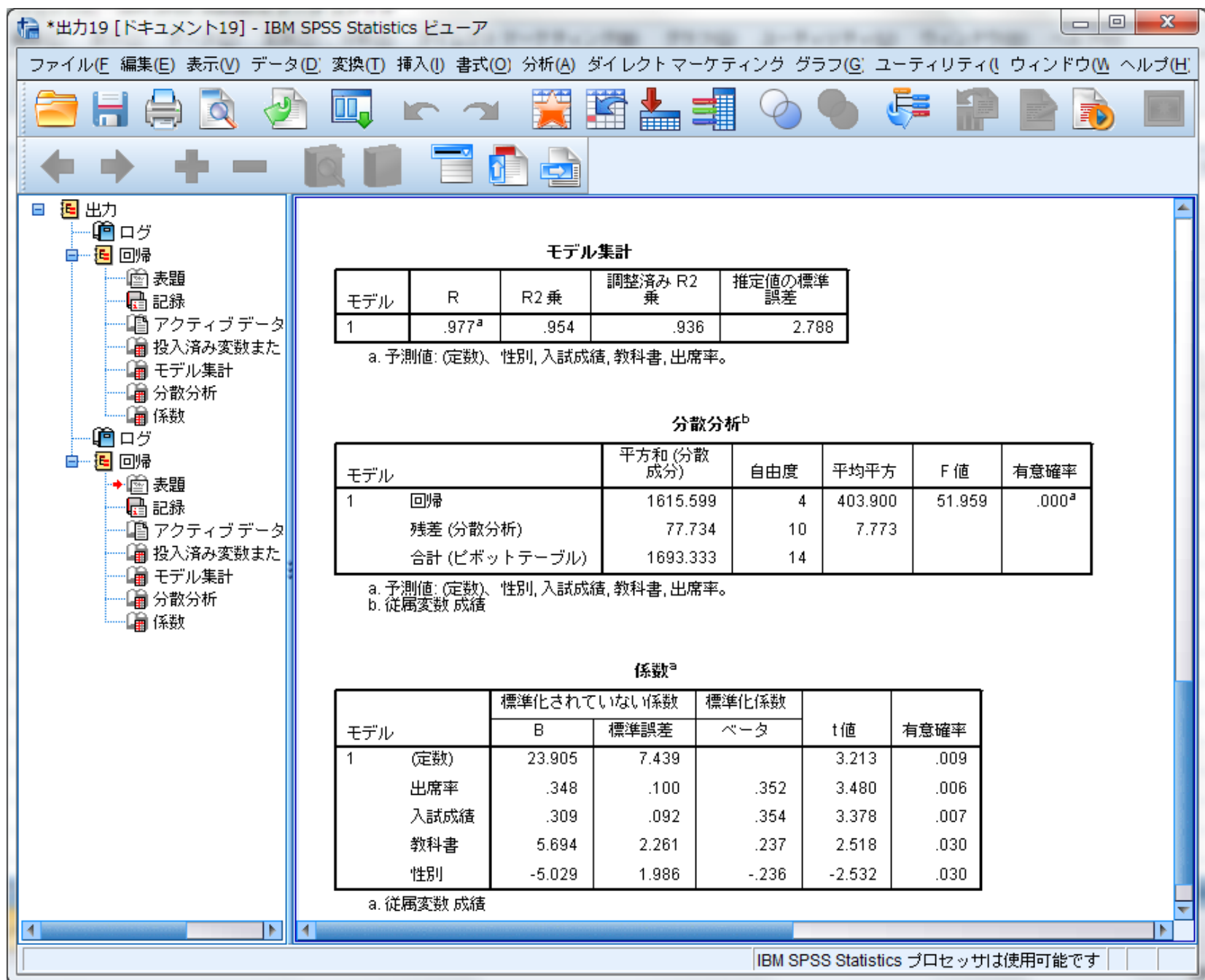
ファイル 編集 表示 データ 変換 分析 ダイレクト マーケテ グラフ ユーティリティ ウィンドウ ヘルプ

16: 性別 表示: 5 個 (5 変数中)

	成績	出席率	入試成績	教科書	性別	var
1	65	75	70	0	1	
2	70	65	70	1	1	
3	90	90	85	1	0	
4	70	65	80	1	1	
5	75	75	80	1	1	
6	85	90	85	1	0	
7	90	90	90	1	0	
8	85	75	80	1	0	
9	75	80	75	1	0	
10	60	55	70	0	1	
11	55	65	40	0	1	
12	70	80	75	0	1	
13	75	70	70	1	0	
14	90	90	95	1	0	
15	80	85	80	1	1	
16						
17						

データ ビュー (D) 変数 ビュー (V)

IBM SPSS Statistics プロセッサは使用可能です



$$Y = 23.91 + 0.35X_1 + 0.31X_2 + 5.69X_3 - 5.03X_4$$

(3.21) (3.48) (3.34) (2.52) (-2.53)

自由度修正済み決定係数 $\text{adj } R^2 = 0.94$

F値 = 51.96

得られた結果をどのように解釈するのか？

① 係数の大きさを比べる

(どの独立変数を操作したほうが従属変数への効果が大きいのか？)

② ?