

行ったダイエットの数を ポイント(スコア)制にして効果を比べる



それでは、ポイント制 (Score) にしてみてもう
でしょうか？ たとえば、それぞれを行うと 1 点
加点する。水を多く摂った日は 1 点、水を多く摂っ
た日とウォーキングをした日は 2 点、全部した日
は 4 点といった具合です。

ID	Water	Walking	Cal	Oil	Score	Weight
1	1	0	1	0	2-3	52
2	0	1	1	0	2	55
3	1	1	0	1	3	50
4	0	1	1	1	3-4	48
5	0	1	1	0	2	53
6	0	0	0	1	1-0	65
7	1	0	1	1	3-1	58
8	1	0	1	1	3	52
9	0	1	1	0	2	59
10	1	1	0	1	3	53
11	1	0	0	0	1-2	56

合計ポイントとして計算したダイエットスコアと体重の相関を調べ
る散布図を描いてみましょう。

- 1日目
- 2日目
- 3日目
- 4日目
- 5日目
- 6日目
- 7日目
- 8日目
- 9日目
- 10日目

1	目
2	目
3	目
4	目
5	目
6	目
7	目
8	目
9	目
10	目
11	目
12	目
13	目
14	目
15	目
16	目
17	目
18	目
19	目
20	目
21	目
22	目
23	目
24	目
25	目
26	目
27	目
28	目
29	目
30	目
31	目
32	目
33	目
34	目
35	目
36	目
37	目
38	目
39	目
40	目
41	目
42	目
43	目
44	目
45	目
46	目
47	目
48	目
49	目
50	目

3.9
~~4.04~~キロ, 2つ行くと ~~8.02~~ ^{7.8}キロ体重が減ることがわかりますね。
 しかし、はたしてこれは正しくダイエット法の効果を表しているのでしょうか？

でも、この4つのダイエット法には、効果のあるものとそうでないものがあるはずですね。また効果のあるものでも、カロリー制限の方が水を多く摂ることより効果が大きいといった場合は、それぞれ同じように1点加点するのは、いかがなものでしょうか？



それぞれのダイエット法の効果に応じて、重みづけした点数計算はできますか？

できますよ。それぞれのダイエット法の効果を「重み」として計算してくれるのが多変量回帰モデルです。



多変量回帰モデルを用いて、 ダイエット法の効果に重みをつける

使用するデータセット **Diet.data**

それでは、多変量解析を行ってみましょう！アウトカムが連続変数の場合は線形回帰モデルを使います。



1 目
2 目
3 目
4 目
5 目
6 目
7 目
8 目
9 目
10 目

	回帰係数 推定値	95%信頼 区間下限	95%信頼 区間上限	標準誤差	t 統計量	P 値
(Intercept)	64.69	55.09	74.30	3.92	16.48	3.18e-06
Cal	-4.03	-10.63	2.57	2.70	-1.49	1.86e-01
Oil	-0.78	-7.05	5.48	2.56	-0.31	7.70e-01
Walking	-6.45	-13.09	0.19	2.71	-2.38	5.51e-02
Water	-6.51	-13.58	0.57	2.89	-2.25	6.53e-02

ここで P 値を見てみると、カロリー制限 (Cal) の P 値は 1.86e-01 となっています。これは 1.86×10^{-1} という意味です。つまり P 値 = 0.186 です。ウォーキング (Walking) の P 値は 5.51e-02, つまり P 値 = 0.0551 です。P 値が 0.05 以上であることから、それぞれのダイエット法の効果には統計的有意差は出ていません。

統計的有意差を調べるときに用いる P 値ですが、それぞれの回帰係数推定値がゼロかどうかを検定しています。決して母集団の体重の減少が 4.02 キロであるかどうかを断定しているわけではありません。

仮に母集団を日本人の 20 代の女性とします。ランダムに選ばれた 20 代の女性 11 人がカロリー制限をすると 4.02 キロ体重が減少しました。この場合の P 値とは、日本の 20 代の女性全員がカロリー制限を行ったときに平均的な体重の減少がゼロの場合 (カロリー制限に効果がないとき) に、その母集団からランダムに集めてきた 11 人の女性で 4.02 キロ以上の体重減少が観察される確率を表します。このデータではカロリー制限の P 値は 0.186, つまり 18.6% の確率でカロリー制限に減量効果がなくても、たまたま 4.02 キロ (かそれ以上) の差が出る可能性があるという意味になり、母集団での体重減少はゼロでないと言い切ることはできない、つまり「統計的有意差には至らなかった」と結論づけられます。このように複数の項目を用いて 1 つのアウトカムを解析する場合に一番よく用いられるのが多変量回帰モデルです。本書では多変量回帰モデルについて詳しく学習します。