

2つの変数の関係

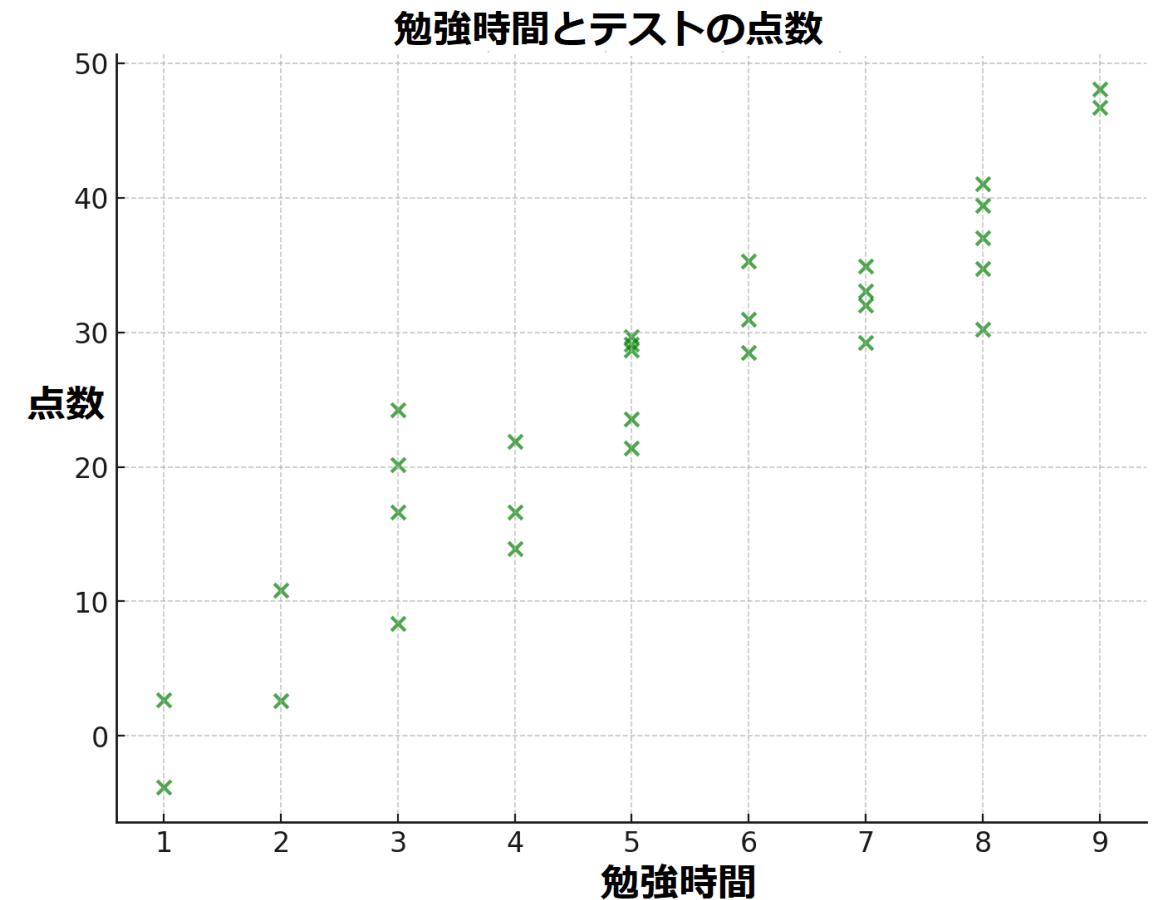
散布図と相関係数を用いて考察する



散布図

2つのデータの関係を座標を用いて表した図

横軸に勉強時間
縦軸に点数
勉強時間とテストの点数
の関係を
図のように示したもの
を散布図と言います。

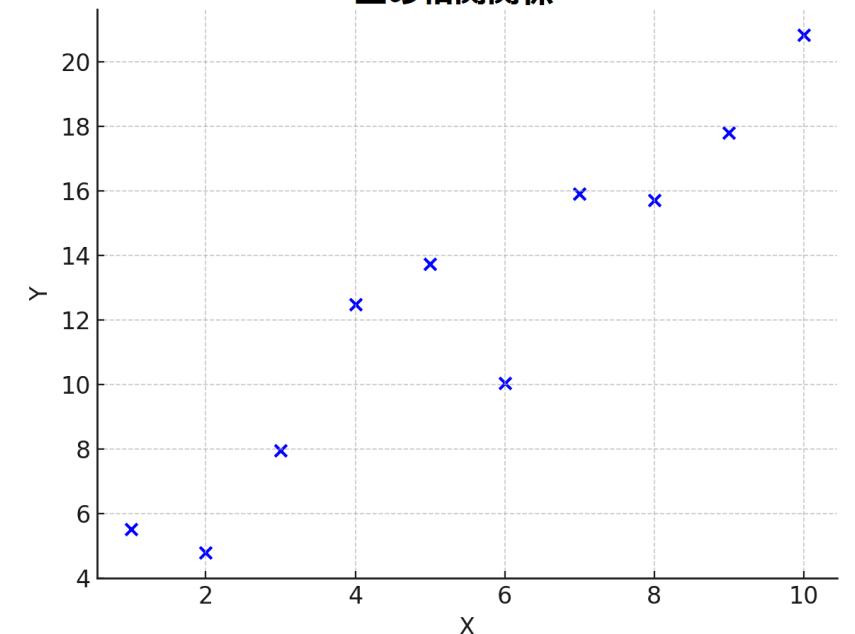


正の相関関係

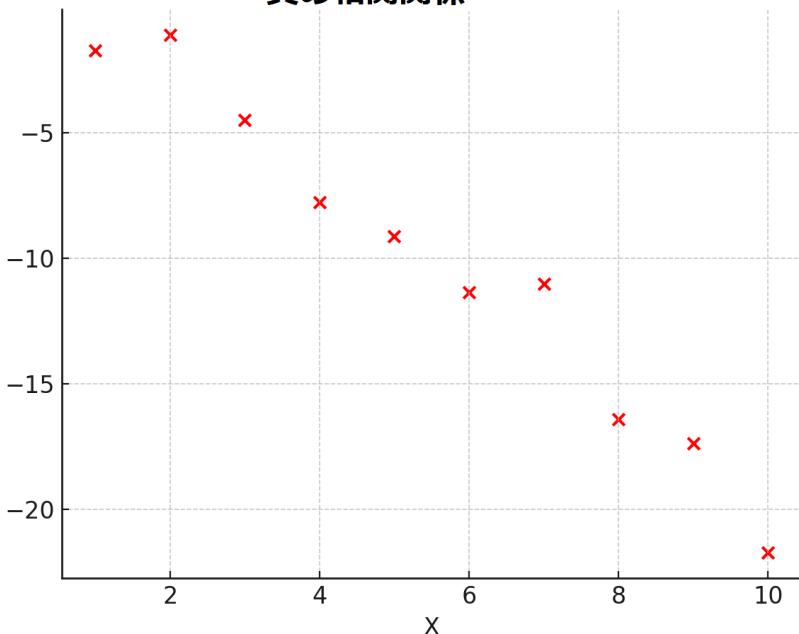
負の相関関係

相関関係はない

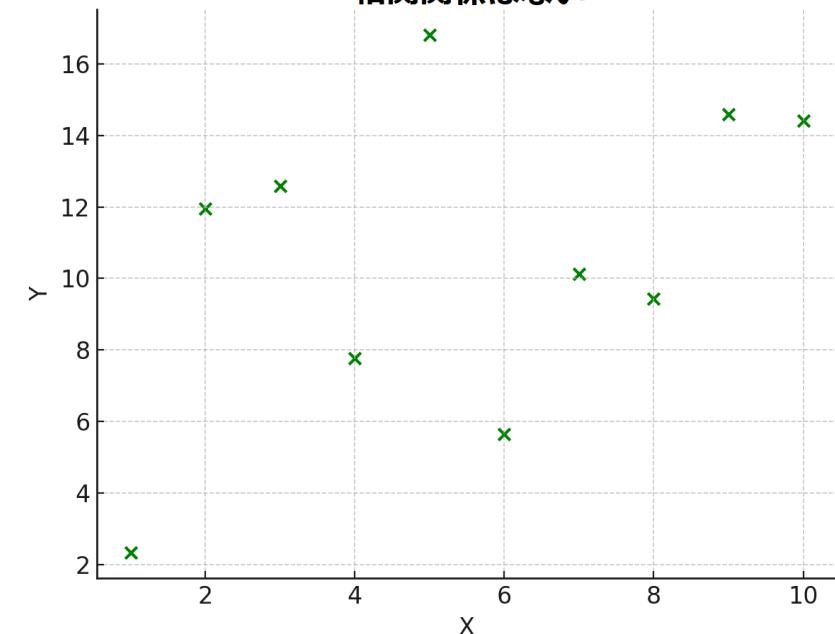
正の相関関係



負の相関関係



相関関係はない



相関係数

- 2つのデータがどれくらい関係があるか数値で示すことができる
- 2つのデータの偏差の積の平均をそれぞれの標準偏差で割った値
- データXとYの相関係数

$$\text{(相関係数)} = \frac{\{(X\text{の偏差}) \times (Y\text{の偏差})\}\text{の平均値}}{(X\text{の標準偏差}) \times (Y\text{の標準偏差})}$$



偏差 $X - (X\text{の平均})$

標準偏差 $s = \sqrt{\frac{1}{n} \{ (x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2 \}}$

相関係数 エクセルで関数機能を使うと

- セルに 「=CORREL(変量1、変量2)」



	A	B	C	D	E	F
1	X	Y				
2	7	13				
3	4	9				
4	9	16	XとYの相関係数	0.953880347		
5	3	7				
6	5	9				
7	3	5				
8	7	12				
9	5	9				
10	9	20				
11	7	13				
12						
13						

例題 散布図と相関係数



ここに20人の学生のデータがあります。

勉強時間、睡眠時間、友達と一緒に勉強した時間とテストの点数です。

勉強時間、睡眠時間、友達と一緒に勉強した時間は、それぞれテストの点数に相関があるか散布図を作成し考えてみましょう。

また、もっとも相関が高いのはどの時間となるか？

散布図で判断が付かない場合は相関係数を求めて判断してください。

学生	勉強時間(時間)	睡眠時間(時間)	友達と一緒に勉強した時間(時間)	テストの点数
1	1	8	0	52
2	3	7	2	56
3	2	6	1	54
4	5	7	3	60
5	4	6	2	58
6	6	5	3	62
7	7	4	4	64
8	8	5	5	68
9	9	3	4	70
10	10	4	5	72
11	4	6	2	58
12	5	5	2	60
13	6	7	3	64
14	7	6	3	66
15	8	5	4	68
16	9	4	4	70
17	10	3	5	72
18	2	6	1	55
19	3	7	2	57
20	4	6	2	59

別資料のデータを参照

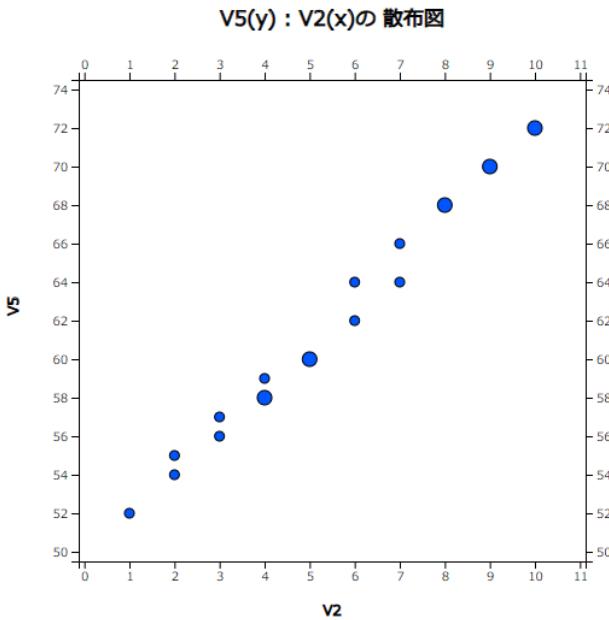
散布図 作成ツール eStat <http://www.estat.me/estat/eStat/indexH.html>

データからグラフ作成が簡単にできます。
1度にグラフで使えるデータは2種類までです。

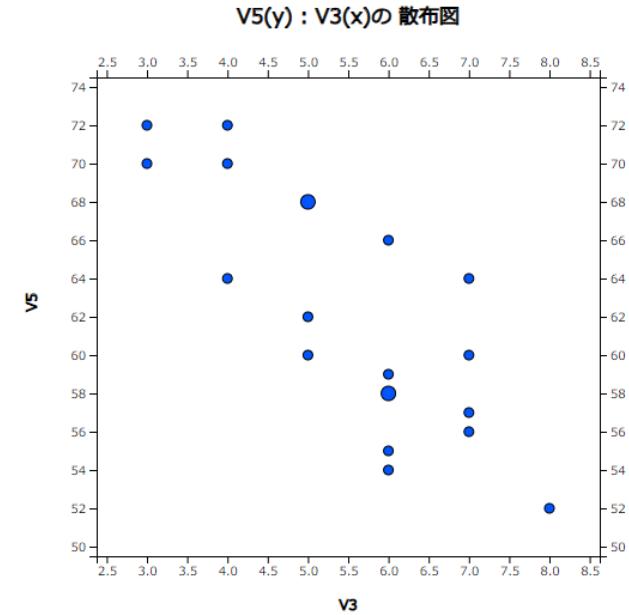
The screenshot shows the eStat software interface. At the top is a toolbar with icons for file operations (Home, CSV, www, JSON, H, json, Print) and various chart types (Bar, Line, Pie, Scatter, Box Plot, Histogram, etc.). To the right of the toolbar is a language selection dropdown set to '日本語'. The main window is divided into three vertical sections. The left section is a CSV file editor for 'Untitled.csv' showing a table with columns V1 through V6 and 16 rows. The middle section is a large workspace with a toolbar above it containing icons for file operations (Save, Print, etc.) and chart types. The right section is currently empty.

例題 解答例

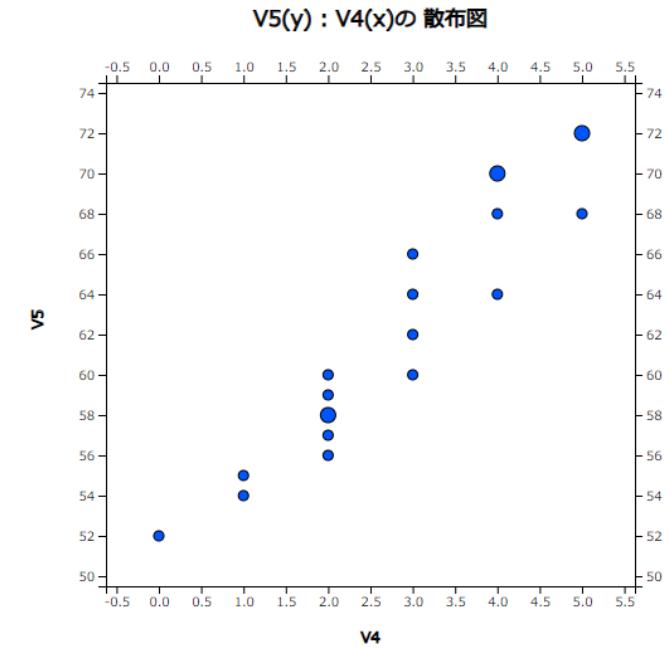
- テストの点数とそれぞれの時間の散布図
- 縦軸がテストの点数 横軸がそれぞれの時間



勉強時間との散布図

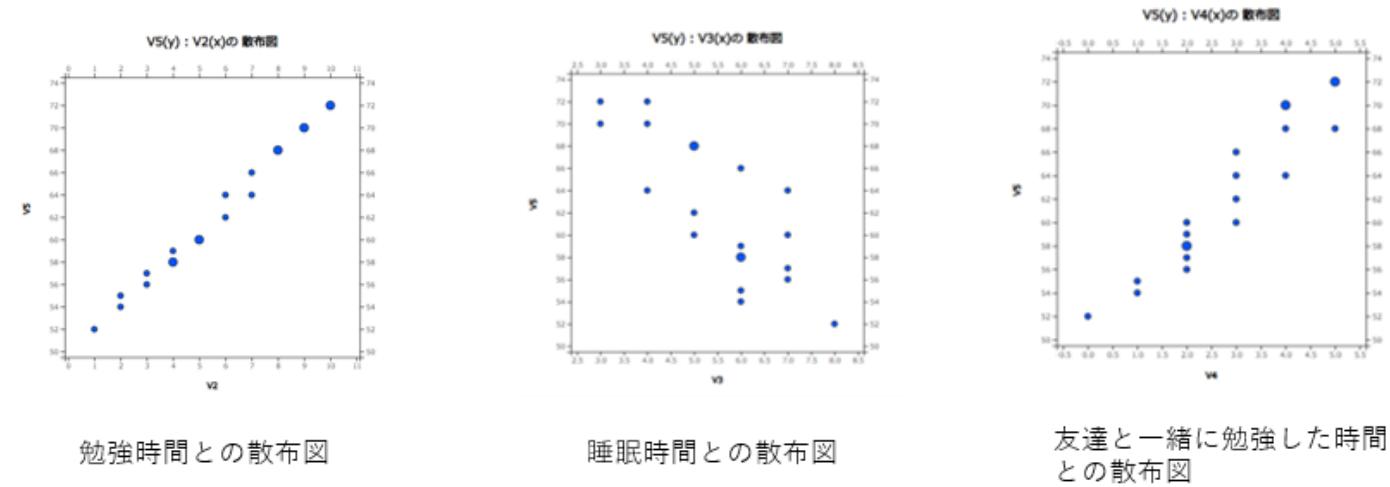


睡眠時間との散布図



友達と一緒に勉強した時間との散布図

それぞれの比較



- ・睡眠時間は負の相関だがやや弱い相関を示している。
- ・勉強時間、友達と一緒に勉強した時間がそれぞれ正の相関で、勉強時間の方が相関は強そうに見える

⇒相関係数を求め比較すると

勉強時間 (時間)	睡眠時間 (時間)	友達と一緒に勉強した時間 (時間)
0.994435	-0.79766	0.948074

勉強時間が0.046361だけ値が大きく相関が強いことが確認できた

課題

パリオリンピック 陸上競技では女子やり投げで、北口選手が日本選手で初めて金メダルに輝く快挙があった。

ここに高校の陸上競技部でやり投げをしている選手30人のやり投げと体力テストのデータがある。どの種目がやり投げの結果と相関があり、どんなことに重点を置いた練習をすれば効率的かデータから考察せよ。

(散布図や相関係数を用いて考察すること)

やり投げの記録 (m)	1500m走 (秒)	50m走 (秒)	ハンドボール投げ (m)	立ち幅跳び (cm)	握力 (kg)	長座体前屈 (cm)	上体起こし (回/分)	反復横跳び (回/20秒)
59.97	478.17	8.49	30.47	195.73	37.23	33.13	23.7	28.06
53.62	429.95	9.46	31.65	142.67	33.9	25.71	27.44	31.02
61.48	454.48	8.81	27.23	212.5	33.49	24.65	27.98	41.94
70.23	450.81	7.15	33.48	182.65	43.3	32.41	27.58	50.8
52.66	464.68	8.66	24.37	169.72	33.06	28.88	12.65	19.59
52.66	525.99	8.11	19.01	201.79	28.02	33.57	23.39	28.25
70.79	410.57	5.91	36.88	192.56	51.8	32.37	13.51	37.72
62.67	508.11	6.46	32.64	176.68	39.97	29.64	16.43	31.07
50.31	538.61	9.58	25.18	152.92	24.23	25.77	10.24	47.84
60.43	452.37	8.56	29.04	171.22	39.54	22.43	18.95	40.31
50.37	476.37	9.56	18.11	120.1	25.35	27.77	38.26	17.61
50.34	493.5	8.39	23.07	152.39	34.14	34.28	-3.57	39.38
57.42	473.79	8.55	27	151.01	40.24	31.07	24.09	55.67
35.87	565.55	11.26	13.92	117.08	17.42	23.77	-5.37	31.85
37.75	593.36	15.07	18.07	94.86	27.47	30.87	6.61	7.46
49.38	524.08	9.3	26.71	179.14	31.69	31.93	25.7	24.79
44.87	534.34	10.94	31.87	118.94	31.03	25.58	14.1	39.59
58.14	435.73	8.67	29.94	167.98	44.37	30.77	6.66	27.81
45.92	506.01	10.72	24.25	154.03	26.33	30.29	6.62	31.99
40.88	589.37	13.81	20.07	98.02	20.76	24.29	19.06	32.27
69.66	411.64	6.29	25.24	213.53	37.35	31.79	13.59	32.53
52.74	500.59	9.09	26.24	184.36	27.56	32.8	17.99	31.05
55.68	497.59	7.39	28.14	134.89	33.02	35.42	17.16	1
40.75	518.65	12.37	32.69	125.94	26.16	35.27	5.71	14.21
49.56	470.83	10.9	23.82	153.88	31.12	23.11	36.31	27.21
56.11	447.62	9.28	29.56	183.97	37.8	25.31	23.17	21.19
43.49	551.22	10.39	21.57	105.73	26.16	32.58	-7.2	42.42
58.76	474.24	7.92	23.54	149.87	42.52	32.57	19.49	20.95
48.99	494.1	10.73	30.21	157.41	28.07	32.58	8.08	24.99
52.08	462.41	9.07	29.8	162.18	44.85	49.26	24.15	32.56

別資料のデータを参照

解答例

やり投げの記録と他の種目との相関係数

1500m走: -0.813 (※逆相関が強い)

50m走: -0.875 (※逆相関が強い)

ハンドボール投げ: 0.583 (正の相関)

立ち幅跳び: 0.779 (正の相関)

握力: 0.810 (正の相関が強い)

長座体前屈: 0.109 (相関がほとんどない)

上体起こし: 0.399 (正の相関)

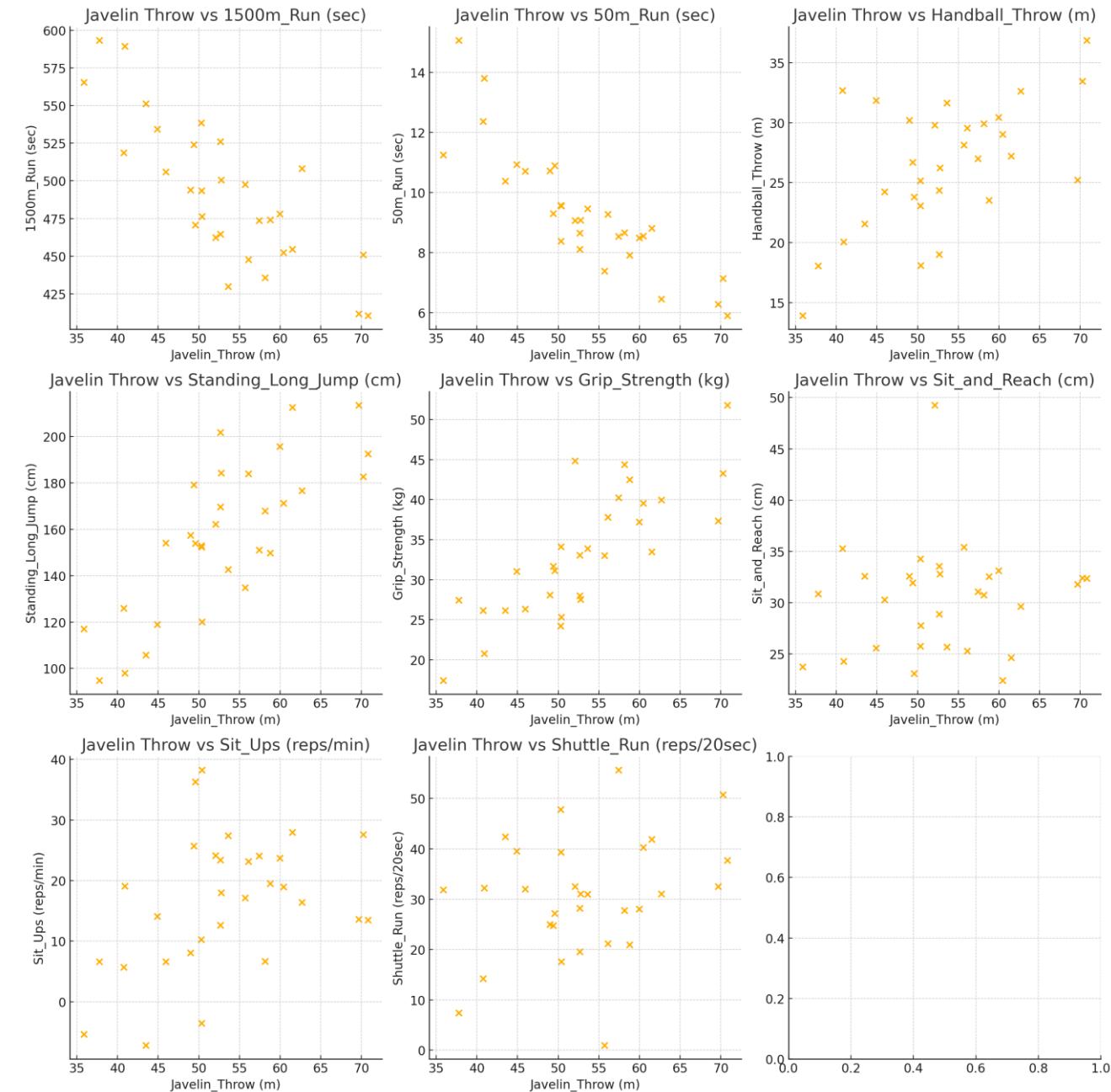
反復横跳び: 0.299 (やや正の相関)

※タイムとしての値が小さい程早く走っている為、実際には正の相関で考える必要がある

考察

1500m・50m走、立ち幅跳びの相関が強いことより脚力強化のトレーニングが有効

1500m、50m走の比較から、より短い距離のトレーニングの方が有効と言える。また握力の強化も有効



考察 その他の例

- やり投げの記録 上位2選手についてはほとんどの種目においてもトップレベルであり、偏りが少ない。トップレベルを維持するには全ての種目で高いレベルのトレーニングが必要
- 長座体前屈: 0.109 より柔軟性を上げるトレーニングは少なくともよい
- 散布図から走るトレーニングの有効性は明らかであり、走れば走るほど記録が上がる可能性がある

データについて

- この教材で使われている全てのデータはChatGPT4.0を使用したものである。（表・画像含む）
- 一部の散布図も同様にChatGPT4.0を使用している。
- eStat <http://www.estat.me/estat/eStat/indexH.html>
- 例題・課題で使われているデータ：別資料参照
(スライド5・7・12の元データ)
- 数学Ⅰ データの分析を習った上での授業が好ましい