

定量的なデータの分析 ～統計的仮説検定と相関

情報 I 第49回授業

06情報通信ネットワークとデータベース

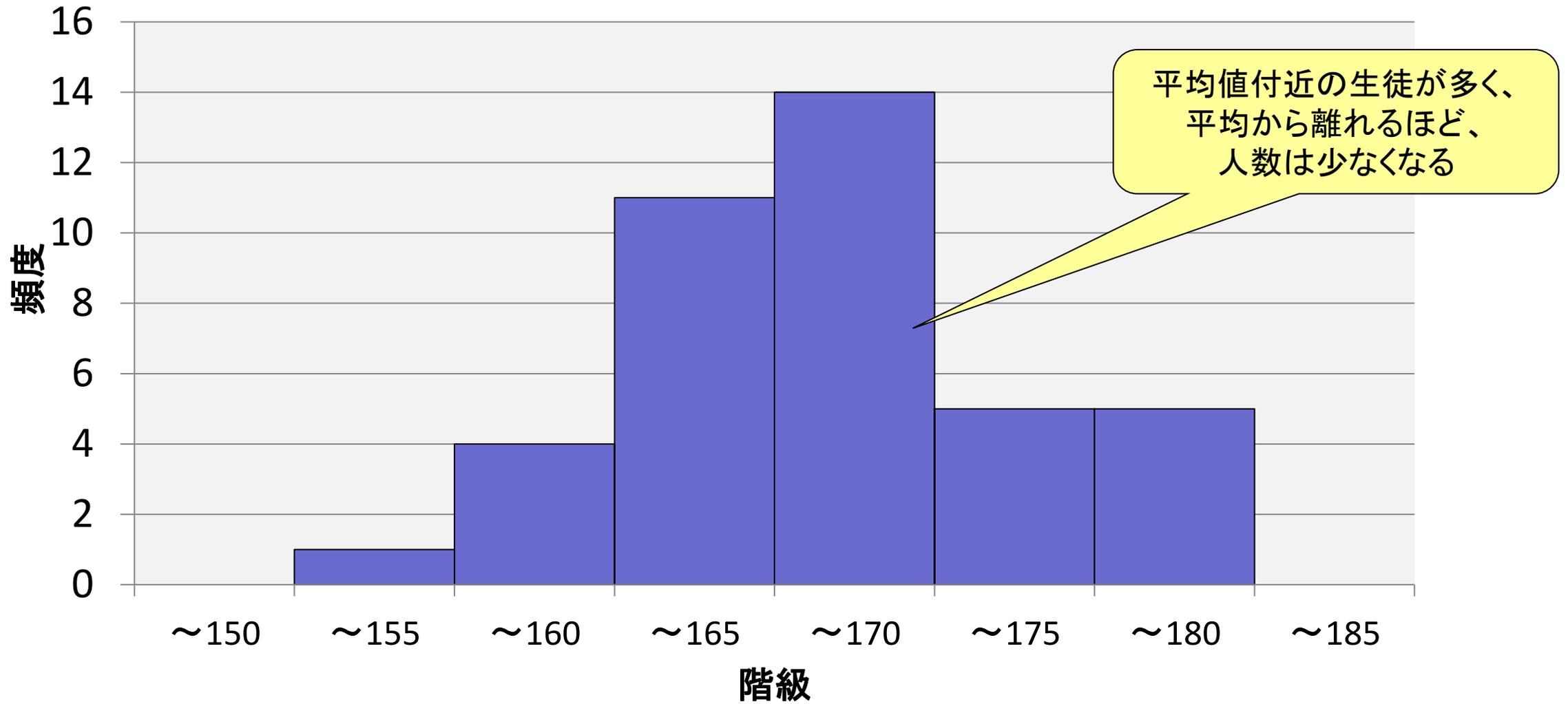
対応データ 23exp48.xls(前回)

今日のテーマ

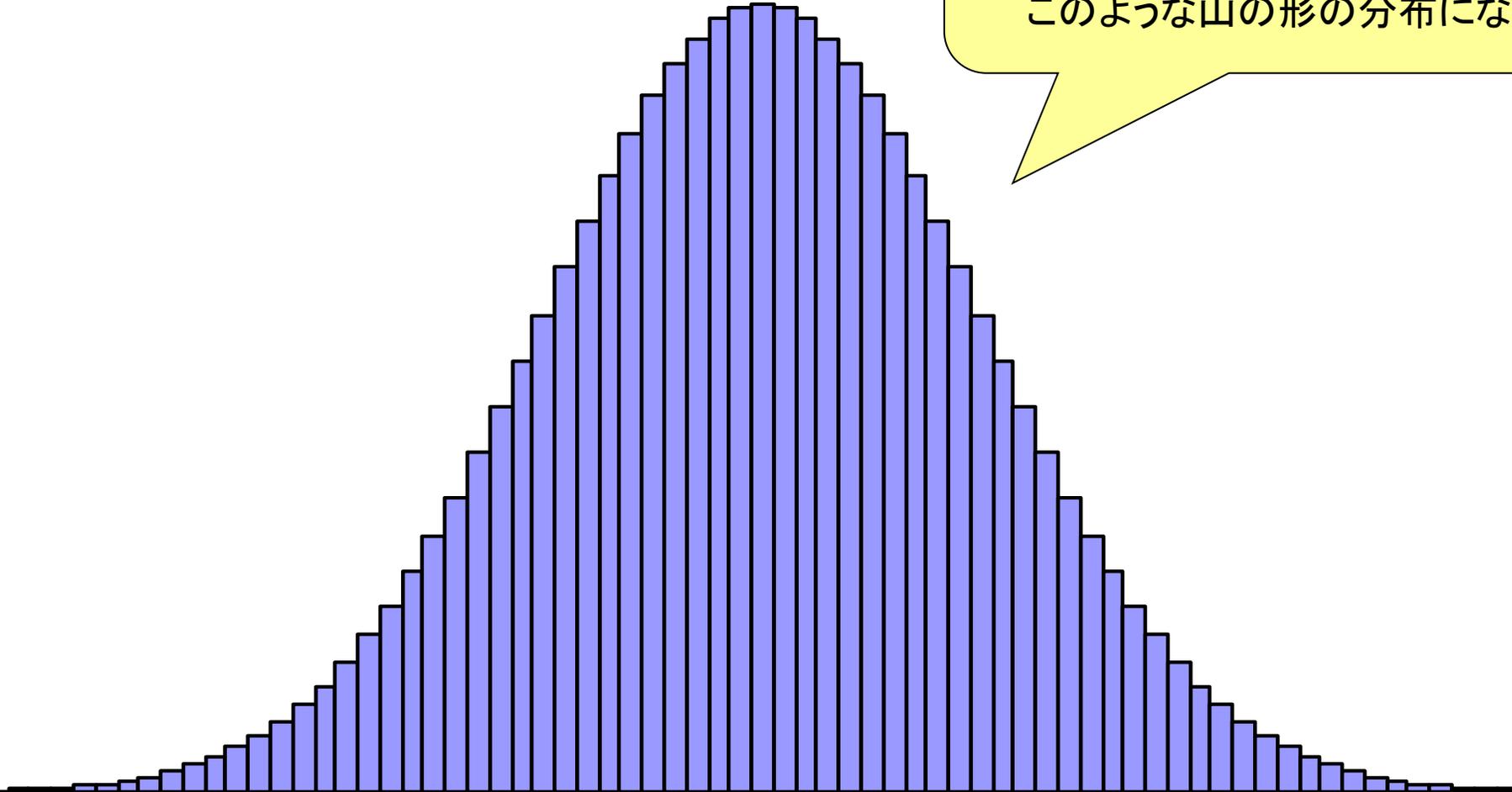
- 本当かどうかを確かめる
- 手がかりをつかむ
- わかりやすく示す

1 「仮説検定」について

A組



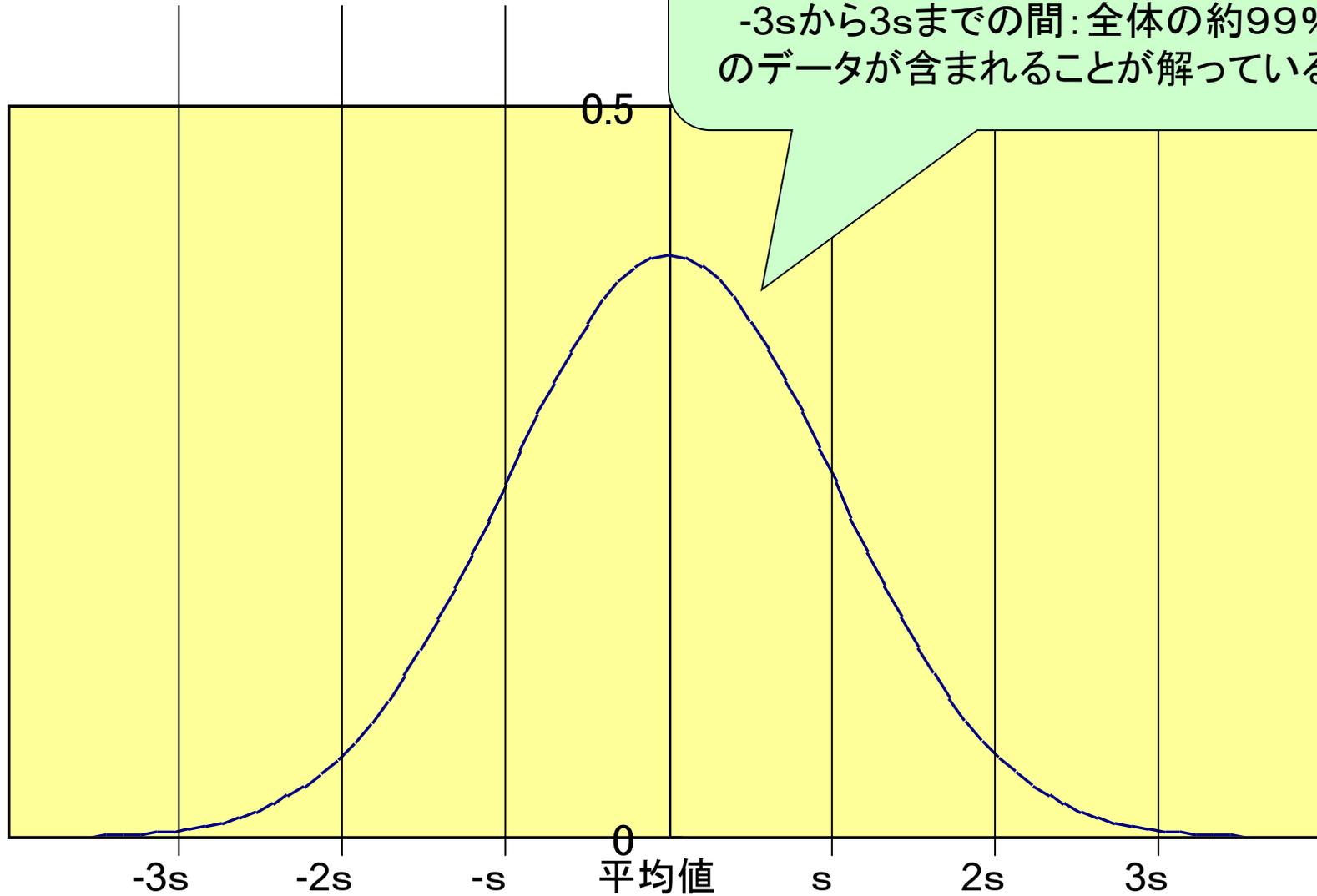
標本(サンプル)の数を増やし、
階級(それぞれの区間)の幅(差)を
限りなく小さくしていくと、
このような山の形の分布になる



正規分布

- 平均値の周辺が最も度数が多く、
- 平均値から離れるに従って、度数が少なくなっていくような分布（山のような形）
- 世の中の多くの分布が、ほぼ正規分布のような形になると見なすことができる。
 - 模試の結果、体重、身長、・・・
- 正規分布で「分かっていること」を活用できる

標準偏差が s の正規分布



統計的仮説検定

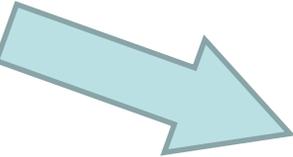
ある出来事に着目した場合、
その起こった出来事の確率が、
一定[有意水準といいます]
(0.05 あるいは 0.01)以下の場合に、
「違いがあるのでは」とする考え方

5回連続で表が出たコイン！

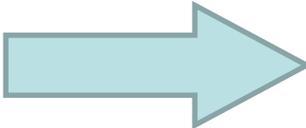
インチキだ！

H₁

偶然だ！



本物なら、そんなに連続しない！



そんなに言うなら「証明」してよ

じゃあ、そのことが起こる確率で決着つけよう！

こんなにめったにおきないことが起こるなんて！



インチキ！

H₁



起きても別に不思議じゃない確率でしょ？

偶然！

H₀

例) 5回連続で表が出たコイン。
「表が出やすい」といえるか？

H_0 : コイン表裏の出やすさは同様

→ 「守り」(=帰無仮説)

H_1 : コインは表が出やすい

→ 「示したいこと」(=対立仮説)

確率を計算し、「基準値」以下ならば、 H_0 は棄却(H_1 を採用)

→ 「示したいこと」が示される(「有意差がある」という)

「検定」と「過誤」

$$(0.5)^5 = 0.03125$$

有意水準 5%

H_0 を棄却 → このコインは表が出やすい

※もちろん「本当は表裏同じ」かもしれない！ ← 第1種の過誤

有意水準 1%

H_1 を棄却 → このコインは表が出やすい

とはい切り切れない

※もちろん「本当は表が出やすい」かもしれない！ ← 第2種の過誤

<注意！>この場合、決して「表裏が同じ」と言い切れるわけではない！

<練習2>

1の目が3回連続して出たサイコロがある。
このサイコロは1の目が出やすいと言えるか。
有意水準1%で検定せよ。

H_0 :このサイコロの目の出やすさは同様
 H_1 :このサイコロは1の目が出やすい

$$1の目が3回連続 \cdots (1/6) \times 3 \doteq 0.00463 < 0.01$$

よって、 H_0 を棄却 → このサイコロは、有意水準1%で
1の目が出やすい！

(発展・参考)

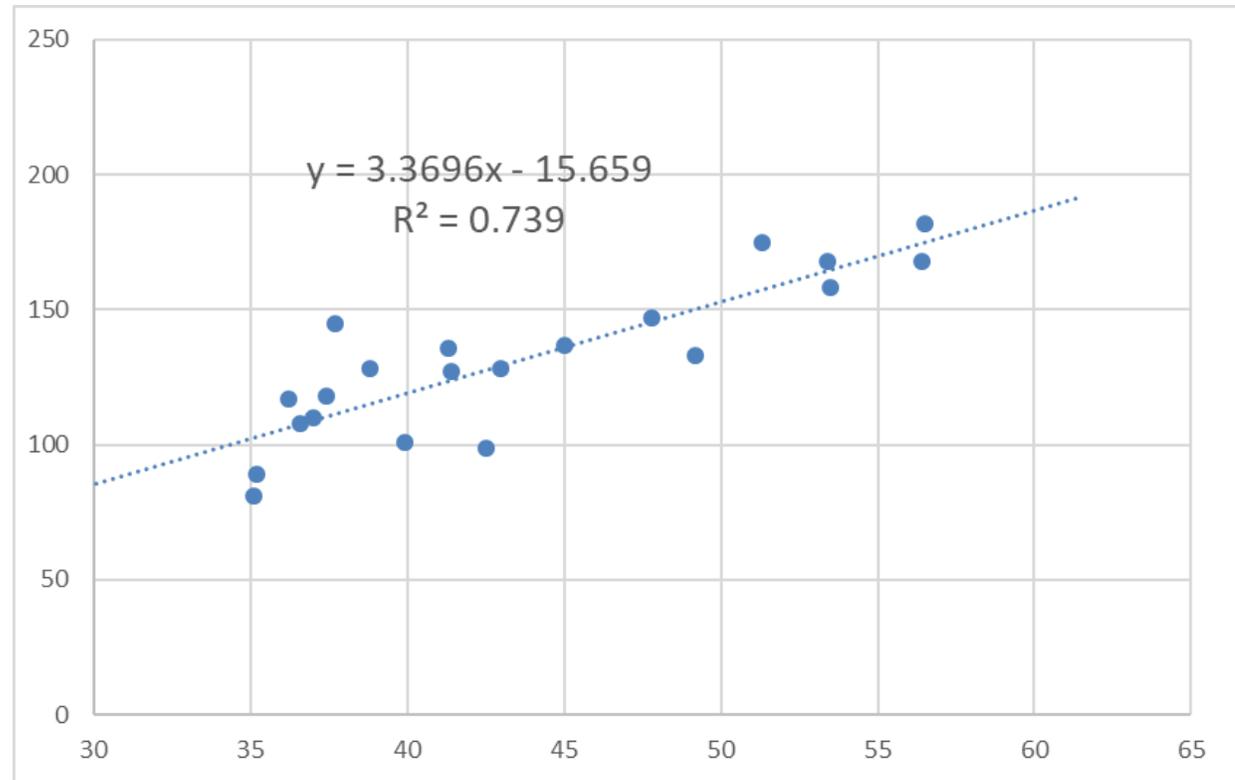
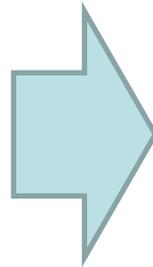
既に知られている分布を活用する

- 分散に「違い」があるかを確認する
 - 「F分布」の活用 → F検定
- 平均に「違い」があるかを確認する
 - 「正規分布」の活用 → Z検定 ($n \geq 30$)
 - 「t分布」の活用 → t検定 ($n < 30$)
- クロス集計表に「違い」があるかを確認する
 - 「 χ^2 (カイニ乗)分布の活用」 → χ^2 検定
 - これらの「表」から得られた確率に基づき、有意差があるかを判断している。

2 相関について

相関を調べる

名前	握力	背筋力
A	38.8	128
B	35.2	89
C	36.2	117
D	56.5	182
E	37.7	145
F	41.3	136
G	37.4	118
H	53.4	168
I	41.4	127
J	35.1	81
K	47.8	147
L	53.5	158
M	36.6	108
N	43	128
O	49.2	133
P	45	137
Q	39.9	101
R	51.3	175
S	56.4	168
T	37	110
U	42.5	99



※この直線を「回帰直線」という

< 練習3 >

- ワークシートにある「握力と背筋力」のデータから、
 - 散布図を作成する
 - 回帰直線を表示させる
 - 回帰直線の方程式を表示させる

相関行列

	身長	体重	座高	握力	上体起こし	長座体前屈	反復横跳び	シャトルラン	50m走	立ち幅跳び	ハンドボール投げ
身長	1.000										
体重	0.382	1.000									
座高	0.756	0.497	1.000								
握力	0.250	0.559	0.315	1.000							
上体起こし	0.066	0.092	-0.029	0.360	1.000						
長座体前屈	0.257	0.235	0.235	0.317	0.309	1.000					
反復横跳び	0.149	0.110	0.093	0.386	0.457	0.477	1.000				
シャトルラン	0.142	-0.090	0.029	0.175	0.341	0.277	0.372	1.000			
50m走	-0.211	-0.098	-0.215	-0.454	-0.329	-0.294	-0.544	-0.553	1.000		
立ち幅跳び	0.359	0.063	0.273	0.412	0.256	0.361	0.483	0.341	-0.674	1.000	
ハンドボール投げ	0.292	0.315	0.278	0.470	0.457	0.408	0.519	0.400	-0.490	0.419	1.000

<練習4>

- 「科学の道具箱」を開き、「高等学校体力測定データ」を確認する
- すでに整形されたデータを元に、相関行列を作ってみる
- 相関行列から相関の高い2系列を選び、散布図と回帰直線を作成する