TI-Nspire CX を使った
新しい学習指導要領に対応した統計の授業と教材

中澤房紀（Naoco Inc./東日本国際大学）

本原稿は，当日配布資料のダイジェスト版です。

☰ ヒストグラム・箱ひげ図

「スーパー立川店の販売会議における鈴木さんの提案」

平日の売り上げデータを分析しました。その結果，
「一人当たりの平均購入金額は 3,030 円」でした。
一人当たりの購入価格を 3,500 円まで引き上げたいのと考え，
「3,500 円以上ご購入のお客様に次回来店時に使える 200 円の商品券プレゼント」というキャンペーンを実施したい。
という提案です。
あなたは，この提案にどう対応しますか。

表 10 月 1 日の販売データ

No	購入額	No.	購入額	No.	購入額	No.	購入額	No.	購入額	No.	購入額
1	749	26	735	51	1,089	76	4,780	101	1,873	126	1,953
2	2,897	27	6,725	52	1,412	77	863	102	10,980	127	5,278
3	6,450	28	1,692	53	798	78	1,156	103	1,574	128	937
4	7,960	29	215	54	7,280	79	9,950	104	623	129	1,272
5	1,962	30	2,012	55	8,900	80	467	105	2,538	130	1,612
6	5,356	31	1,002	56	378	81	1,821	106	5,850		
7	967	32	3,870	57	1,736	82	1,518	107	7,415		
8	1,337	33	747	58	538	83	603	108	5,198		
9	1,658	34	6,940	59	5,780	84	2,437	109	892		
10	127	35	8,760	60	4,480	85	5,750	110	1,215		
11	6,600	36	1,712	61	802	86	4,977	111	1,592		
12	7,980	37	298	62	398	87	875	112	652		
13	1,984	38	2,089	63	1,777	88	1,162	113	2,612		
14	5,421	39	532	64	1,092	89	7,400	114	5,900		
15	502	40	5,518	65	1,457	90	10,780	115	7,688		
16	972	41	1,077	66	528	91	477	116	498		
17	1,389	42	1,393	67	2,265	92	1,862	117	1,921		
18	2,901	43	3,980	68	5,728	93	1,539	118	5,245		
19	1,684	44	759	69	7,380	94	612	119	923		
20	176	45	6,980	70	9,800	95	2,489	120	1,263		
21	8,650	46	8,870	71	425	96	5,860	121	669		
22	5,484	47	328	72	1,783	97	5,007	122	2,784		
23	528	48	2,187	73	1,463	98	889	123	5,996		
24	1,391	49	545	74	546	99	1,176	124	7,758		
25	3,780	50	5,659	75	2,397	100	478	125	499		

平均というひとつの代表値でデータを語るとき，その他のすべての情報が失われています。よって，売上データを分布からも見てみます。

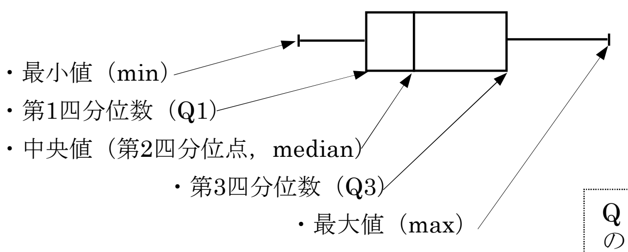
１）箱ひげ図とヒストグラムでデータを分析する

ここでは，グラフ電卓 TI-Nspire を使用して表のデータを分析します。その方法として，「箱ひげ図」，「代表値の計算」，「ヒストグラム」を用います。

箱ひげ図

箱ひげ図は，ばらつきのあるデータをわかりやすく表現するための統計学的グラフの１つです。また，以下の重要な要約統計量を読み取ることができます。

箱ひげ図

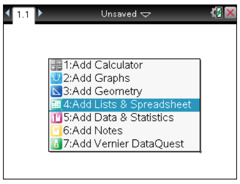


Q は，４分の１を表す Quarter の Q

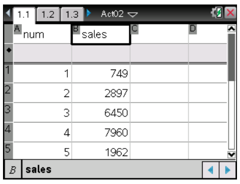
表のデータを TI-Nspire に入力し，分析します。

1 データの入力

「List & Spreadsheet」のページでデータを入力します。



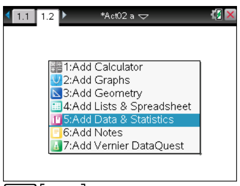
ctrl +page で「List & Spreadsheet」のページを追加します。



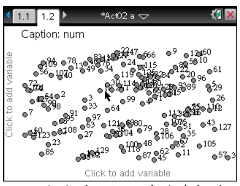
列の名前を入力し，それぞれのデータを入力します。

2 箱ひげ図で調べる最小値・最大値，中央値と四分位数

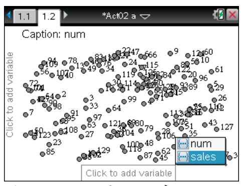
「Data & Statistics」のページでグラフ化します。



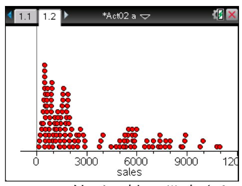
ctrl +page で「Data & Statistics」のページを追加します。



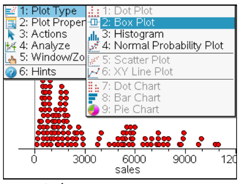
このようなページが追加されます。画面下の「Click to add variable」にカーソルを持っていきます。



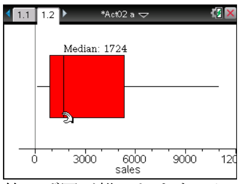
そこで，クリックするか enter を押すと，変数が表示されるので「sales」を選択します。



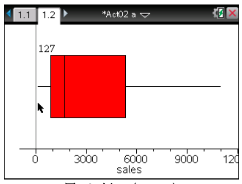
sales の値が x 軸に設定され，ヒストグラムのようにデータが並びます。



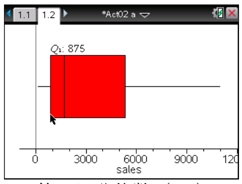
menu を押して，
「1:Plot Type」▸でプルダウンメニューから「2:Box Plot」を選択します。



箱ひげ図が描かれます。カーソルを動かすことで，統計値が表示されます。



最小値（min）

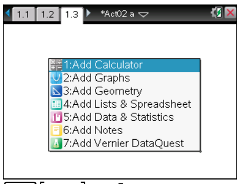


第 1 四分位数（Q1）

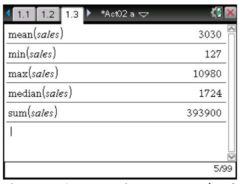
カーソルを左右に移動させて最小値，第1四分位点，中央値，第3四分位点，最大値を調べましょう。得られたデータをまとめます。

- 最小値（min） .127
- 第1四分位数（Q1） 875
- 中央値（median） 1,724
- 第3四分位数（Q3）5,356
- 最大値（max） 10,980

3 代表値を計算で求める



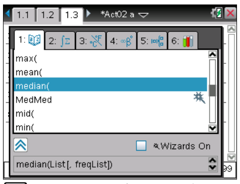
ctrl +page で「Calculator」のページを追加します。



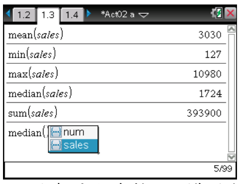
上のように入力して enter 押すと計算されます。

←平均 mean
←最小値 min
←最大値 max
←中央値 median
←合計 sum

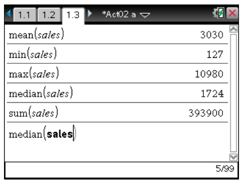
＜関数をカタログから選択する＞



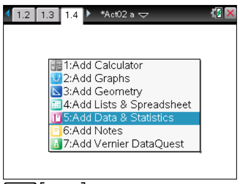
📖 でカタログから関数を選択します。頭 1 文字の英字キィを押すとそこまでジャンプします。enter 押すと上のように関数が入力されます。



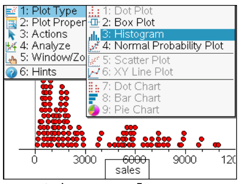
var を押すと変数の一覧が表示されます。そこから必要なものを選択します。



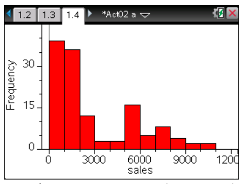
4 ヒストグラムの作成



ctrl +page で
「Data & Statistics」のページを追加し，ここまでは「箱ひげ図」と同じです。

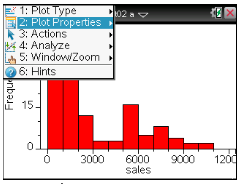


menu を押して，「1:Plot Type」▸でプルダウンメニューから「3:Histgram」を選択します。

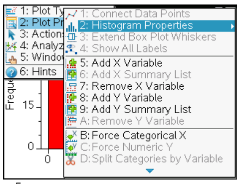


enter 押すとヒストグラムが描かれます。
（階級の幅は自動的に 1000 になっています。）

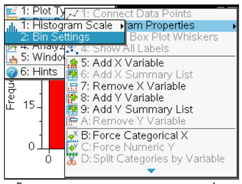
❑ 階級の幅を変更する。ここでは 500 にします。
WINDOW の設定を変更する。度数の y 軸の最大を 30 にします。



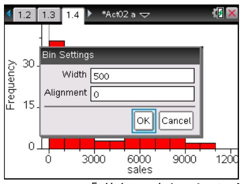
menu を押して
「2:Plot Properties」で▶。



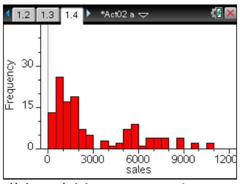
「2:Histogram Properties」で▶。



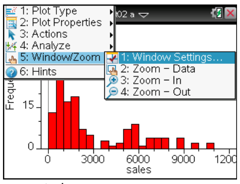
「2:Bin Settings」で enter 押します。



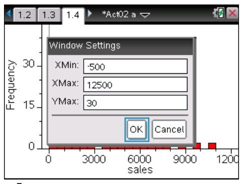
Width で「階級の幅」を入力します。「OK」にして enter 押します。



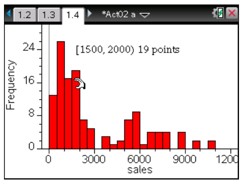
階級の幅を 500 にしたヒストグラムが描かれます。



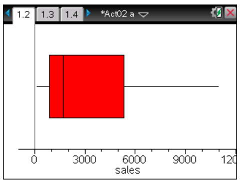
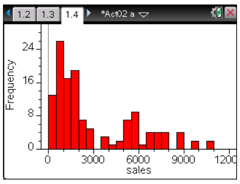
menu を押して，
「5:Windows/Zoom」▶で
プルダウンメニューから
「1:Window Setting」を選択します。



「1:Plot Type」▶で
それぞれの値を入力して
「OK」にして enter 押します。



❏ 分析結果をまとめます。

箱ひげ図	ヒストグラム
	
最小値（min）	127
第 1 四分位数（Q1）	875
中央値（median）	1,724
第 3 四分位数（Q3）	5,356
最大値（max）	10,980
四分位偏差（Q3-Q1）	4,481
平均値	3,030

階級	度数	
0-500	13	
500-1000	26	←最頻値
1000-1500	17	
1500-2000	19	
2000-2500	7	
2500-3000	5	
3000-3500	0	
3500-4000	3	
4000-4500	1	
4500-5000	2	
5000-5500	7	
5500-6000	9	
6000-6500	1	
6500-7000	4	
7000-7500	4	
7500-8000	4	
8000-8500	0	
8500-9000	4	
9000-9500	10	
9500-10000	2	
10000-10500	0	
10500-11000	2	

これらの分析データから言えることを以下にまとめます。

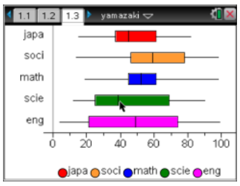
✓ このスーパーの平均客単価は3,030円である。
✓ 分布を見ると25%のお客さんは875円以下であり，
　半分のお客さんは1,724円以下である。
✓ 平均が3,030円となっているのは，5,000円以上購入する
　お客さんが25%いることによる。

この結果を踏まえて，
あなたは，先の鈴木さんの提案に対してどのような対応をされますか。

❑ヒストグラムと箱ひげ図及び平均値でデータの分布を考える

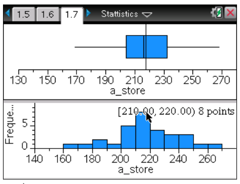
・複数のデータを比較する。

３つのお店のデータを比較

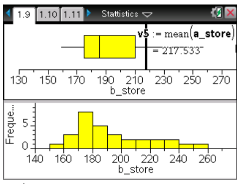


国語，社会，数学，理科，英語の点数の分布の比較

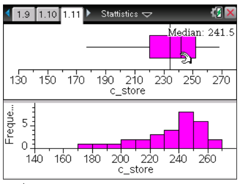
・最頻値，平均値，中央値とデータの分布



A）



B）



C）

A）平均値≒中央値≒最頻値であるデータの分布は，富士山のような正規分布に近いものになります。

B）平均値>中央値>最頻値であるデータの分布は，左に山がある分布となっています。

C）平均値<中央値<最頻値であるデータの分布は，右に山がある分布となっています。

☰ 分散と標準偏差

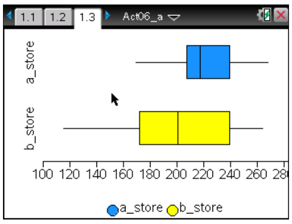
「２つのコーヒーショップの本日の売上額は，ともに 17 万円」

あなたは，２つのコーヒーショップを経営しています。
今日の売上額をそれぞれの店長さんに電話で聞きました。
　A店：本日は 17 万円です。
　B店：本日は 17 万円です。
あなたは，この売上の数字をどのように判断しますか。
右表にあるそれぞれのお店の 20 日間の売上データをもとに，その根拠を示して説明してください。

単位（千円）

日付	A店	B店
1	249	239
2	187	264
3	218	245
4	220	158
5	214	116
6	246	239
7	200	263
8	210	186
9	217	228
10	205	188
11	258	210
12	233	229
13	258	115
14	202	211
15	213	184
16	268	159
17	169	191
18	229	249
19	216	187
20	228	115

１）大きくデータをとらえる。<A 店と B 店の統計量>

	A 店	B 店
平均	222	198.8
最小値（Min）	169	115
第１四分位数（Q1）	207.5	171.5
中央値（Med）	217.5	200.5
第３四分位数（Q3）	239.5	239
最大値（Max）	268	264
四分位範囲	32	67.5

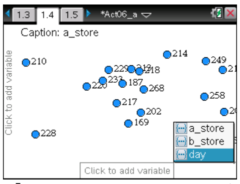
箱ひげ図が２つの店舗の傾向を視覚的に表しています。どのようなことが言えるかまとめてみます。

- 日々の売上額のばらつきは A 店が B 店に比べて少ない。
- A 店の最小値である約 17 万円は，B 店の第 1 四分位数の値に近い。B 店では概ね 4 分の 1 の営業日で A 店より売上額が少ない。
- 第 3 四分位数の値は，両店とも約 24 万円であり，概ね 4 分の 1 の営業日がこの数値以上の売上額である。
- 最大値は両店とも概ね 26 万円である。

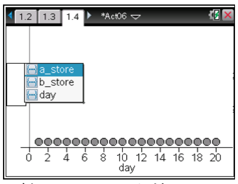
２）データがどれだけ平均値の周りに散らばっているかを知る。分散と標準偏差

1 散布図を描いて，そこに平均値を乗せる

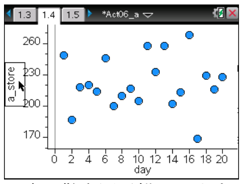
ctrl +page で「Data & Statistics」のページを追加します。



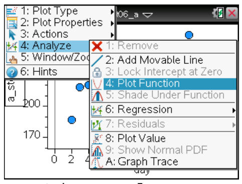
「Click to add variable」をクリックするか enter を押すと，変数が表示されるので「day」を選択し enter を押します。



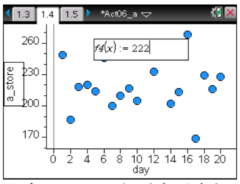
y 軸にカーソルを持っていき，クリックするか enter を押すと，変数が表示されるので「a_store」を選択し enter を押します。



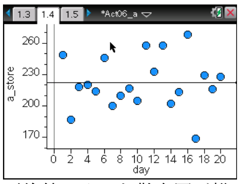
A 店の散布図が描かれます。



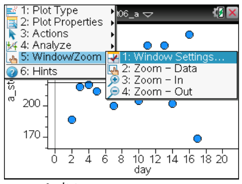
menu を押して，「4:Analyze▸」で「4:Plot Function」を選択します。



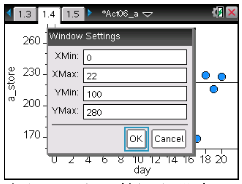
入力フィールドが表示されるので式を入力し，enter を押します。



平均値の入った散布図が描かれます。

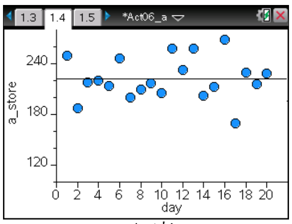


menu を押して，「5:Window/Zoom▸」で「1:Windows Settings…」を選択します。

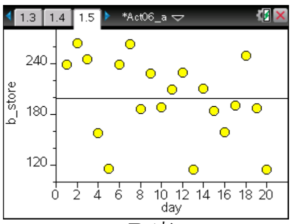


上記にように範囲を設定し「OK」で enter を押します。

❑ A 店と B 店の散布図



A 店



B 店

散布図を見ると平均値の周りのばらつきは，A 店より B 店のほうが大きいことがわかります。

データのばらつきを数値的に表わすことはできないか？

平均値に対するばらつきですから，平均値と比較してどれだけデータが離れているかを考えます。

分散

分散（variance）はデータの散らばり具合を表す量で，対象となるデータの平均を $\bar{x}$ とすると，

分散　$\sigma^2 = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(x_i - \bar{x})^2$　と定義されます。

標準偏差

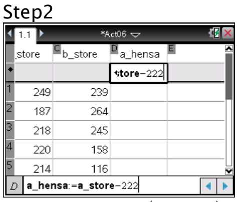
分散は，元のデータを 2 乗しているので，元のデータあるいは平均値と直接比較することができません。そこで平方根をとって単位をそろえます。標準偏差のことを英語で、standard deviation といいます。頭文字をとって、SD と表記します。

ここでは TI-Nspire を操作しながら，標準偏差を求める手順を示します。

Step1．平均を求める。
Step2．偏差（個々のデータから平均を引いた値）を求める。
Step3．偏差の 2 乗を求める。
Step4．分散（偏差の 2 乗の平均）を求める。
Step5．標準偏差（分散のルート）を求める。

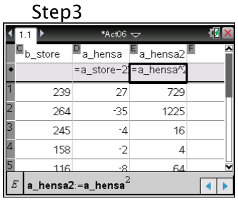
Step1．既に求めています。A 店は 222，B 店は 198.8 です。

Step2

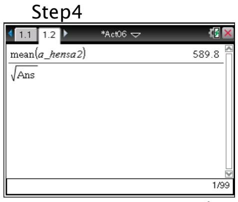


=a_store－mean(a_store)

Step3



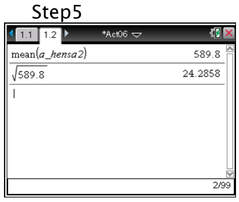
Step4



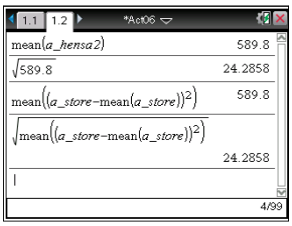
Ans は，ctrl ans で入力します

変数は，文字列を入力する方法と，var で呼び出し選択する方法があります。

Step5



計算の仕方を整理して立式すると以下のように計算できます。
標準偏差の計算がイメージ可できていると立式は容易です。



a_store の各データから「a_store の平均」を引き，それらのデータの 2 乗を計算し，その平均をとり，そのルートを計

☰ 散布図と相関係数

> **「江崎グリコ株式会社の有価証券報告書の内容です。」**
>
> (1)天候による影響
> 当社グループが展開している事業の中には，菓子・アイスクリーム・ヨーグルト・飲料等，気温の高低や晴雨という天候状況によって消費者の購買行動が影響を受けやすい商品があり，春秋の低温，猛暑，多雨をはじめとする天候不順の場合は当社グループの業績に悪影響を及ぼす可能性があります。

アイスクリームの製造量が天候に左右されとのこと，気温との関係で具体的な数字をもとに調べてみます。

表　アイスクリームの生産量と東京の平均気温

	2008		2009	
月	平均気温	生産量（Kl）	平均気温	生産量（Kl）
month	temp_2008	pro_2008	temp_2009	pro_2009
1	5.9	8,706	6.8	7,183
2	5.5	8,153	7.8	9,640
3	10.7	12,114	10.0	10,207
4	14.7	11,180	15.7	11,208
5	18.5	11,814	20.1	11,178
6	21.3	12,192	22.5	11,403
7	27.0	11,923	26.3	12,510
8	26.8	13,223	26.6	12,573
9	24.4	11,857	23.0	11,395
10	19.4	12,082	19.0	10,534
11	13.1	10,966	13.5	9,217
12	9.8	9,825	9.0	9,131

列名

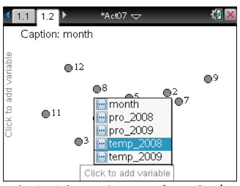
・乳製品の生産量 アイスクリーム（乳脂肪分８%以上のもの）
・牛乳乳製品統計(農林水産省統計情報部)
・東京の気温：気象庁

視覚的にデータをとらえるために散布図を描いてみましょう。

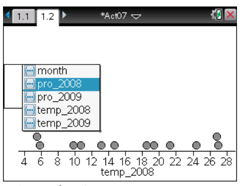
１）　散布図を描く

x 軸に平均気温，y 軸に生産量をとった散布図を描きます。

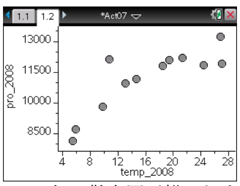
ctrl +page で「Data & Statistics」のページを追加します。



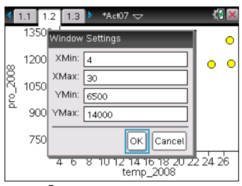
画面下をクリックすると変数一覧が表示されるので，「temp_2008」を選択して enter を押します。



画面左をクリックして「pro_2008」を選択して enter を押します。

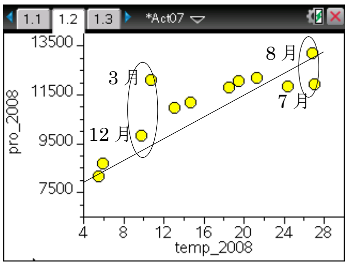


2008 年の散布図が描かれます。

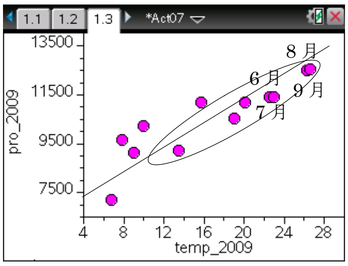


menu「5:Window/Zoom▸」「1:Window Setting..」で，X と Y の範囲を修正します。

同じように 2009 年の散布図も描きます。



2008 年の散布図



2009 年の散布図

散布図からわかることをまとめます

- 2008 年も 2009 年も気温が高くなる概ね生産量が増えている。
- 2008 年の散布図は, 中ほどのデータは一直線上に並んでいるが, 7 月と 8 月, 3 月と 12 月は, 気温が概ね同じにもかかわらず生産量は随分違う。
- 2009 年の散布図は概ね一直線上に並んでいるように見える。
  7 月と 8 月，6 月と 9 月は気温も生産量もほとんど同じである。

気温が決まれば生産量が決まるというような 1 次関数で表現できるわけではないが，どちらの散布図も気温と生産量との間に関連がありそうです。2 つを比べてみると, 2009 年のほうがシャープで, 2008 年の散布図はそれに比べると少し“ばらけている”感じがします。
これは，気温と生産量の関係は 2009 年の方が強くて，2008 年の方が弱かったと言えます。
では,

◇ 関連性が強い，弱いを数字にできないか。

2）相関係数を求める

相関係数

相関係数 r (correlation coefficient)は，気温と生産量，身長と体重，広告費と売上などといった 2 つの変数 x, y の間に，どの程度の関連があるかを測るための指標です。相関係数を計算することによって，x と y の間に，どの程度の直線的な関係があるか（＝データが直線の近くにどのくらい集中しているか）を知ることができます。
相関係数は以下の式で求めます。

$$\text{相関係数 } r = \frac{\text{偏差積の平均}}{(x\text{の標準偏差})\times(y\text{の標準偏差})}$$

Step1．平均気温の偏差と偏差の 2 乗を計算する

Step2．生産量の偏差と偏差の 2 乗を計算する

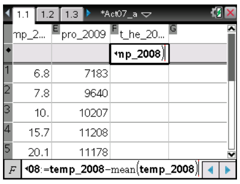
Step3．偏差積を求める

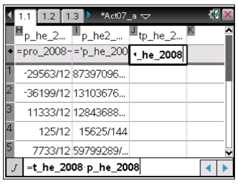
Step4．平均気温と生産量の分散と標準偏差を求める

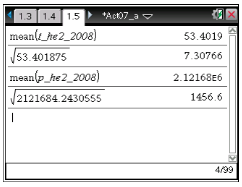
Step5．偏差積の平均を求める

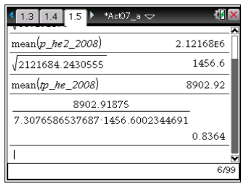
Step6．相関係数を求める

$$\frac{\text{偏差積の平均}}{(\text{気温の標準偏差})\times(\text{生産量の標準偏差})}$$









❑ 結果

2008 年の平均気温と生産量の相関係数は，0.836 となります。この 2 つの間には，「強い正の相関」があります。

±0.7～±1	強い相関がある
±0.4～±0.7	中程度の相関がある
±0.2～±0.4	弱い相関がある
±0～±0.2	ほとんど相関がない

見せかけの相関

生ビールの売り上げとアイスクリームの売り上げの相関は強いと考えられます。これは，両方の変数に気温という変数が共通しているからと考えられます。つまり，

「気温が高いから，生ビールの売り上げが増える」

「気温が高いから，アイスクリームの売り上げが増える」

という因果関係が同時に成立しているので，見かけ上，2 つの間の相関が強くなっています。これを見かけの相関と呼びます。

外れ値

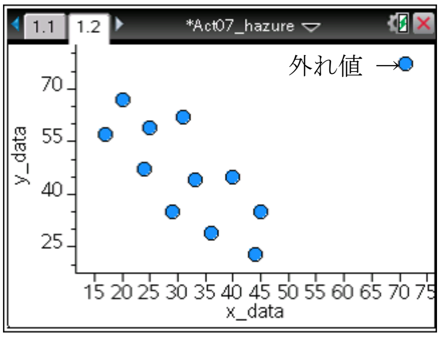
相関の強さを表す相関係数は 2 つの間の関係を見るのに有効なものですが，「外れ値（Outliers)」が大きく相関係数の値に影響を及ぼし本来の関係を見誤る可能性があります。

外れ値とは，全体の分布の中心から極端に外れた値のことです。データの入力ミスであれば修正すればいいのですが，そうでない場合は，外れ値の扱いを慎重に検討しなければいけません。

右の表の散布図を描くと下記のようになります。

表　外れ値

No.	x-data	y-data
1	25	59
2	17	57
3	**71**	**77**
4	29	35
5	33	44
6	20	67
7	40	45
8	45	35
9	44	23
10	36	29
11	31	62
12	24	47



右上に 1 つだけ離れたデータがあります。このようなデータを外れ値と言います。上記のデータの相関係数を求めると 0.054 となります。No.3 の（71,77）の外れ値を除外して相関係数を求めると－0.743 となります。このように，外れ値は相関係数に大きな影響を与えることがあります。データから除外するかどうかは，目的と照らし合わせて慎重な検討が必要です。