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・代表値

図１はある中学校の第１学年の男子生徒 100 人のハンドボール投げの記録である。資料から傾向を読み取ってください。

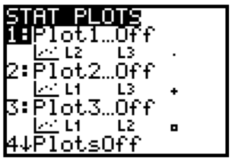
目的に応じて資料を収集し，それを表やグラフに整理し，代表値や資料の散らばりに着目してその資料の傾向を読み取ることができるようにする。計算はテクノロジーで。

ここでは，データの整理にグラフ電卓を使ったときの方法を説明します。

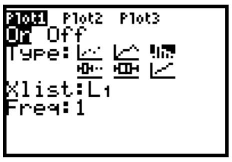
16,	12,	27,	18,	18,	23,	22,	24,	15,	13
26,	12,	24,	24,	15,	10,	18,	15,	18,	18
18,	18,	15,	16,	21,	11,	12,	20,	26,	27
16,	20,	25,	21,	18,	18,	23,	16,	18,	24
16,	18,	14,	18,	14,	14,	18,	15,	14,	18
23,	23,	23,	14,	14,	21,	21,	27,	25,	23
20,	22,	27,	18,	18,	14,	18,	18,	27,	24
15,	25,	15,	24,	23,	21,	25,	25,	15,	16
24,	11,	25,	23,	13,	13,	20,	15,	20,	26
18,	20,	25,	22,	23,	23,	21,	22,	16,	22

図１　[出典：中学校学指導要領解説　p78 のデータ]

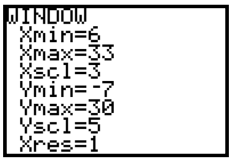
１）ヒストグラムの作成（データの傾向，階級の幅，最頻値）



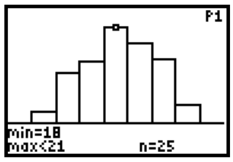
y　，　で「1:Plot1」でÍ　。



OnでÍ　，Type:でヒストグラムを選択して Í　，Xlistでは，データがあるリストを入力。L1であれば y　d を押す。

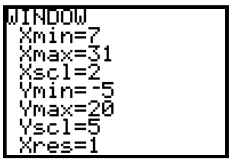


p　を押して数値の範囲（Xmin,Xmax）と階級の幅（Xscl），及び相対度数の範囲（Ymin,Ymax）を指定して　s　を押す。

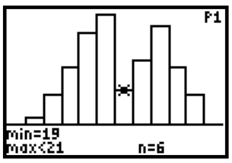


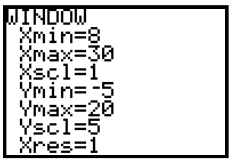
r　を押すことにより，階級と度数が表示されます。

階級の幅を変えてみましょう。階級の幅は，Xscl で設定します。

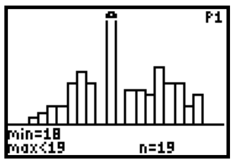


階級の幅＝2





階級の幅＝1



★17mと19mの記録が１つもない！

Activity

階級の幅のよってヒストグラムの形がずいぶん異なります。階級の幅はどのように決めればよいのでしょうか，議論しましょう。

❐ 最頻値

最頻値について理解しましょう。最頻値とは，度数の最も多い階級に対する値です。

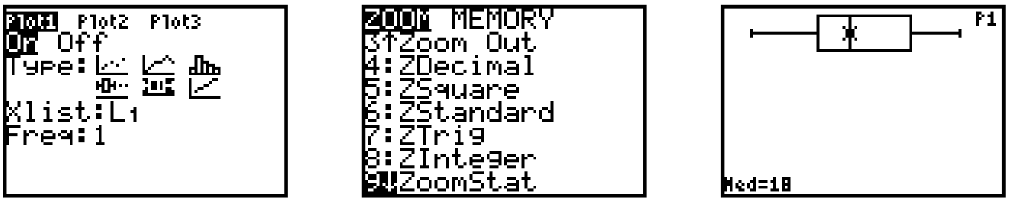
階級の幅＝3のヒスとグラムにおいて,「18以上21未満である」よりその階級の真ん中の値をとって「最頻値は19.5である」する方がよいでしょう。最も望ましい論述はここではふれません。

❐ 相対度数

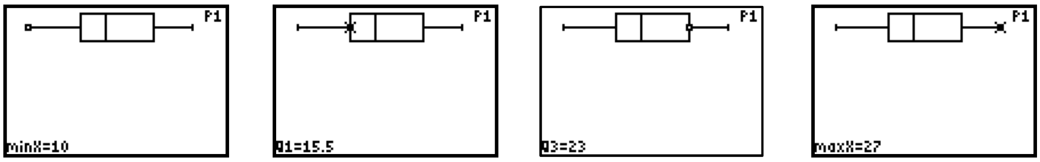
トレースをすることにより，各階級の度数が表示されます。それをもとに相対度数の表を作りましょう。また何故，相対度数が必要かと議論しましょう。

2）データの傾向を大きくつかむための代表値。平均値，中央値，四分位数

❐ 箱ひげ図で調べる中央値と四分位数

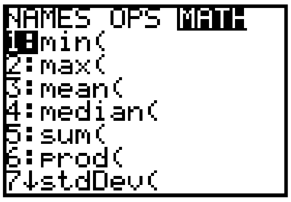


r を押すことにより，中央値が表示されます。

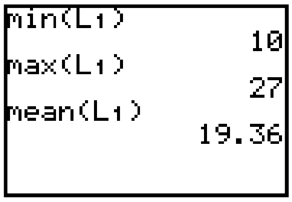


高校で扱う四分位偏差は，第1四分位数と第3四分位数の差として求められます。
この要約統計量は，データがとる中央値から±25%，全体の50%の範囲を示す幅として理解でき，データの散らばりを定量的に把握する手段となります。

❐ 代表値などを計算で求める。



←最小値
←最大値
←平均
←中央値
←合計



y 9 でカーソルで~MATHを選択します。

必要なものを選択してÍ を押します。

求めるべきリストを入力し¤ を付けてÍ を押します。

Activity

20人の生徒がいます。平均点が60点になるように20人分のデータを作り階級の幅を10点としてそのヒストグラム作りなさい。(10点刻みで)

＊生徒が作るデータで多いものが，①100点が10人，20点が10人。②60点が20人。
どちらのクラスで授業したいですか？

Activity

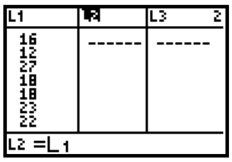
20人の生徒がいます。平均点が60点で最頻値が50点，中央値が70点になるように20人分のデータを作り階級の幅を10点としてそのヒストグラム作りなさい。(10点刻みで)

＊条件となる代表値などの増やすとデータのばらつきが小さくなってきます。

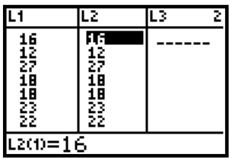
＊データを作ることによって代表値や最頻値についての知識を定着させることもできます。

３）データのコピーや並べ替え，部分的な合計を出す。

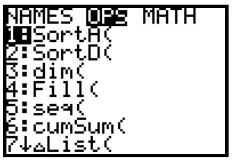
代表値や分布だけではなく，データを加工して議論するときの根拠を持ちましょう。
＊コピー（L1をL2にコピーする）/「並べ替え」 と「総和，部分総和」



＝L1を y d で入力して Í を押します。

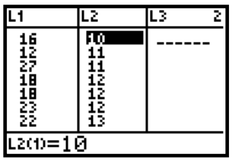


L1の内容がコピーされます。＝以降で計算式が使えます。

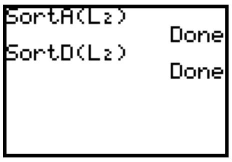


y 9 カーソル~で本画面を表示し「1:SortA」を選択してÍ 。

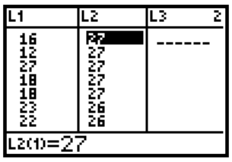
L2を小さい順に並べ替えるので y e ¤ を付けて Í 。



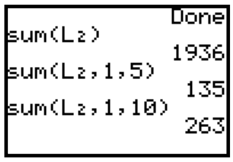
リストを表示すると小さい順に並べ替えられています。



L2を大きい順に並べ替えます。



リストを表示すると大きい順に並べ替えられています。



ここでは，L2の合計，L2の上から5つの合計，同様に上から10個の合計を求めています。

上記の機能を使って，

Activity

Ａ，Ｂ，Ｃの３クラスで対抗リレーをします。各クラス全員の100mのタイムがデータとしてあります。代表選手の数を5人にしたらどのクラスが勝か，10人にしたらどのクラスが勝か，データを加工して得た結果をもとに根拠を持って説明しましょう。

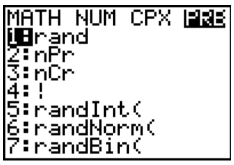
４）標本調査

母集団から無作為抽出により標本を抽出。乱数を利用することにより無作為抽出が可能になる。

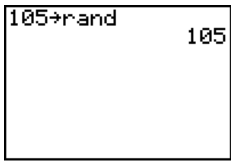
a) 乱数発生の前準備

ここで，乱数の「種」を設定します。
種を別々にしないと乱数機能を使ってもそれぞれの電卓で同じ番号が出てしまいます。
種は自分の生年月日の月日を使いましょう。1 月 5 日なら 105. 11 月 6 日なら 1106。

数字を入力し，¿ を押します。



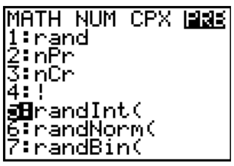
• を押します。右カーソル~でPBRにして，1:randでÍ 。



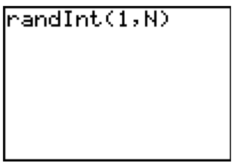
再度，Í を押します。

b) 乱数の発生

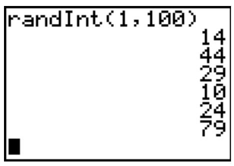
乱数発生の方法を説明します。



• を押します。右カーソル~でPBRにして，5:randInt(でÍ 。



Nのところには，発生させる乱数の最大値を入力します。



Í を押すと乱数が発生します。以後，Í を押します。

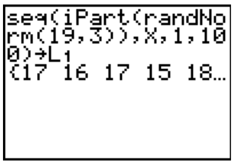
先の「男子生徒 100 人のハンドボール投げの記録」の平均は，19.36m でした。
ここでは，乱数を発生させて何人かのデータを抽出して平均を出してみます。

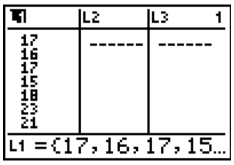
乱数	14	44	29	10	24	79	69	33	25	83	合計	平均
3 人	24	18	26								68	22.7
4 人	24	18	26	13							81	20.3
5 人	24	18	26	13	16						97	19.4
6 人	24	18	26	13	16	15					112	18.7
7 人	24	18	26	13	16	15	27				139	19.9
8 人	24	18	26	13	16	15	27	25			164	20.5
9 人	24	18	26	13	16	15	27	25	21		185	20.6
10 人	24	18	26	13	16	15	27	25	21	25	210	21.0

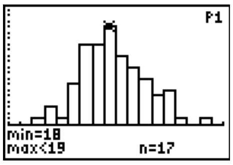
人数を増やしてもいっこうに平均に近づきません。
実データ使うと必ずしも理論通りにならないことがあります。それも議論の対象です。

5）正規分布のデータを発生させる。

Seq(iPart(randNorm(平均，標準偏差)),X,1,100) ¿ d
数列を作成する。Seq y 9 OPS 「5:seq」
整数でつくる。 iPart • NUM 「3:iPart」
正規分布になるように乱数を発生させる randNorm • PBR 「6:randNorm」
平均を 19，標準偏差 3 で (19,3)
変数を X として，100 個のデータを X,1,100
そのデータを L1 に格納する ¿ d







☰ 分散と標準偏差

単位（千円）

あなたは, 2 つのコーピーショップを経営しています。
それぞれの 20 日間の売上データがあります。それぞれのお店の特徴を把握しておきましょう。

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

Activity

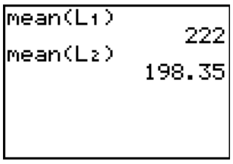
あなたは, 今日の売上をそれぞれの店長さんに電話で聞きました。
A店：本日は 17 万です。B店：本日は 17 万です。
あなたは，この売上の数字をどのように判断しますか。その根拠を示して説明してください。

「少ない」,「超少ない」,「普通じゃない」。
どのような根拠を持ってそのように言うのでしょうか。
基準はないのでしょうか。

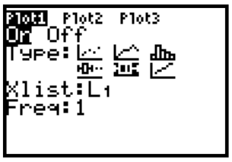
1）大きくデータをとらえる。

❒ 平均と箱ひげ図で調べる中央値と四分位偏差

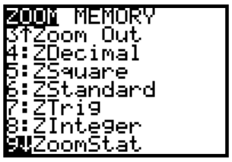
＜A店＞



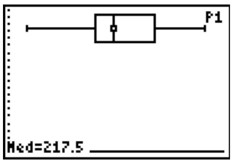
y　9 でカーソルで ~MATHを選択し，「3:mean」でÍ　。



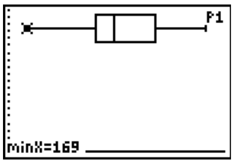
y　,　で「1:Plot1」でÍ　を押し設定。



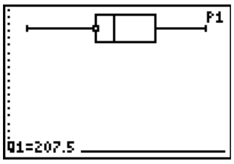
q 「9:ZoomStat」でÍ　を押します。



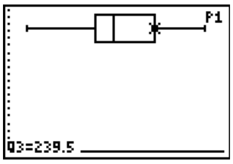
中央値=268



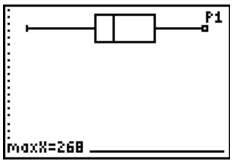
最小値=169



第1四分位数=207.5



第3四分位数=239.5



最大値=268

＜B店＞

同様に求めると中央値=196，最小値=115，第1四分位数=171.5，第3四分位数=239，最大値=264

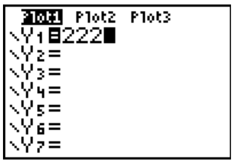
Activity

上記の値を使ってA店，B店の売上傾向についてまとめましょう。
それをもとに，本日の売上 17 万円について論じましょう。

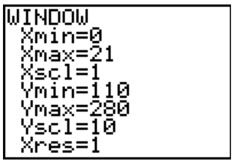
2）データがどれだけ平均値の周りに散らばっているかを知る。分散と標準偏差

データの散らばりを見るために散布図を描いてみましょう。

❒ 散布図

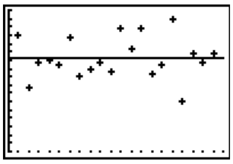


o で平均値を入力します。



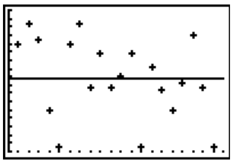
p　の値をA店，B店の最小値，最大値より上記のように設定します。

A店



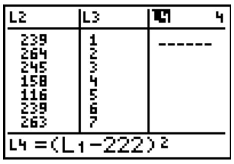
s　を押します。

B店

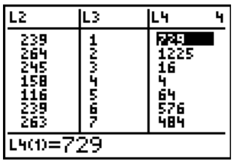


A店と同様に。

❒ 分散と標準偏差

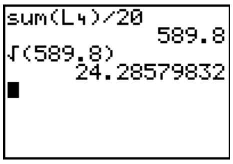


…　「1:Edit」でÍ　を押してデータ編集画面に，上記のように計算式を入力します。



1つのデータを入力したらÍ　を押すことにより計算結果が入力されます。

A店



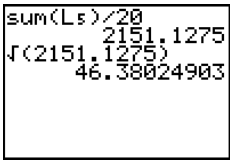
＊分散を計算する。

基本画面で y　9 MATH「5:sum」でÍ を押してL4の総和を計算しそれを個数20で割る。

＊標準偏差を計算する。

その値のルートを計算する。

B店



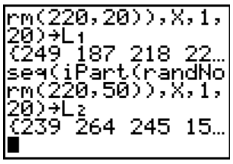
A店と同様に。

Activity

散布図，分散，標準偏差使ってＡ店，Ｂ店の売上傾向についてまとめましょう。
それをもとに，本日の売上 17 万円について論じましょう。

あるデータの集団を代表の値として平均を用いることが多い。**平均＝データの総和÷個数**
平均に違いがなくても，データのちらばり具合は異なることがある。
データのちらばり具合を見るためには，散布図を描く。
データのちらばり具合を示す数値として，分散や標準偏差を用いる。
分散＝（（データー平均値）の二乗）の総和÷個数
標準偏差＝（分散）のルート
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。

＜参考＞
Ａ店，Ｂ店のデータは下記の条件で作成しました。



Ａ店は，平均が 220 で標準偏差が 20　（実際は，平均が 222 で標準偏差が 24）
Ｂ店は，平均が 220 で標準偏差が 50　（実際は，平均が 198 で標準偏差が 46）

☰ 散布図と相関係数

アイスクリームのチェーン店です。立川店が 3 日後にオープンします。どれくらいのお客さんが来るか予測したい。同じような規模の町田店のデータが下記のものです。予測してください。

単位（人）

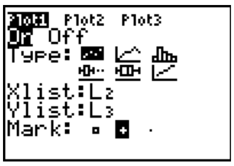
日付	最高気温	客数
1	29	312
2	30	348
3	29	284
4	32	369
5	33	420
6	32	536
7	34	652
8	27	275
9	28	294
10	32	368
11	34	451
12	32	405
13	30	458
14	28	422

Activity

目的のために何をするか。
あなたはどのような関係を知りたい，見たい？
議論しましょう。
日付と最高気温？　日付と客数？　最高気温と客数？

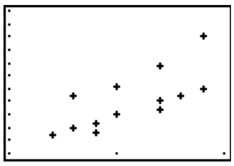
どのような関係を見るか決めたところで，視覚的にデータをとらえるために散布図を描いてみましょう。

１）散布図を描く

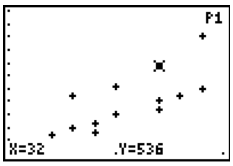


y　，　で「1:Plot1」でÍ　を押し上記の設定。
X軸に最高気温，Y軸に客数。

q　「9:ZoomStat」でÍ　を押します。



散布図は表示されます。



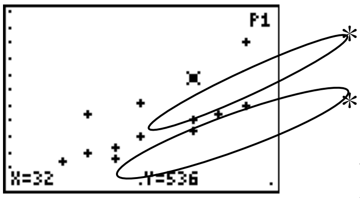
r　を押して，散布図の様子を観察しましょう。

Activity

この散布図からわかることを議論しましょう。

＜想定＞　最高気温が高くなるとお客さんの数も増える。

＊注意深く散布図を見て，どの点が何日目のデータかを見るともう一つの視点が予想されます。



＊最高気温が高くなるとお客さんの数も増える。

＊２つのグループによって形成されていることに気付くでしょうか。上のグループは 6 日，7 日，13 日，14 日のデータであることがトレースにより確認できます。
土曜にと日曜日のようです。

２）最高気温と客数の間に関係があるというが，何かちょっと

平日だけにした立川店とＢ店のデータです。両店の最高気温と客数の関係を見てみましょう。

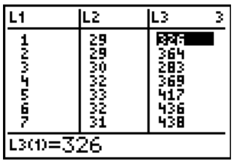
立川店　　単位（人）

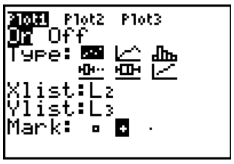
日付	最高気温	客数
1	29	326
2	29	364
3	30	283
4	32	369
5	33	417
6	32	436
7	31	438
8	26	296
9	28	263
10	31	389
L1	L2	L3

町田店　　単位（人）

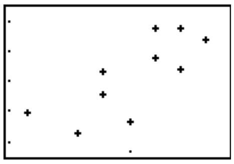
日付	最高気温	客数
1	29	312
2	30	348
3	29	284
4	32	369
5	33	420
6	27	275
7	28	294
8	32	368
9	34	451
10	32	405
	L4	L5

Ａ）散布図を描く

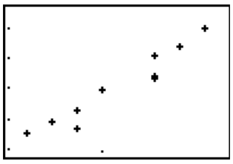




L2,L3で立川店



立川店の散布図



町田店の散布図

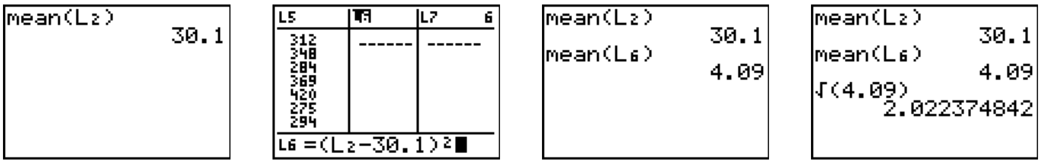
どちらの散布図も右上がりの正の相関ですが，比べてみると，町田店の方がシャープです。
立川店の散布図は，何というか，ばらけている感じです。
最高気温と客数の間の関係が町田店の方が強くて，立川店の方が弱いといえます。

相関が強い，弱いかを数字にできないか。

Ｂ）相関係数を求める

立川店の相関係数を求める

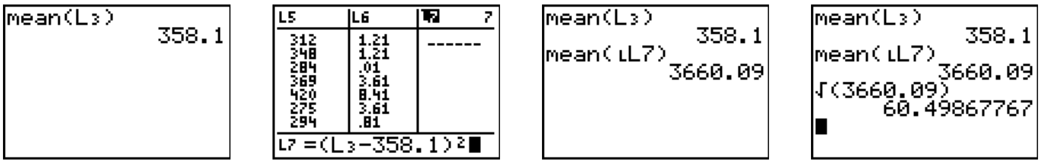
1)　L2 の標準偏差を求める。平均を求める。L6=(L2-平均)²。sum(L6)/10。その値をルート。



y　9　でカーソルで
~MATHを選択し，
「3:mean」でÍ　　を押
します。

標準偏差は，2.02

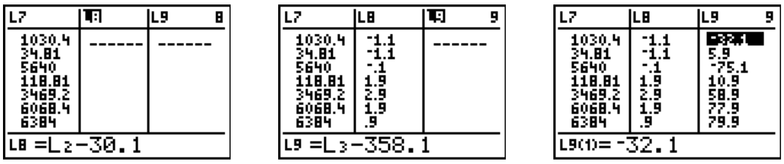
2)　L3 の標準偏差を求める。平均を求める。L7=(L3-平均)²。sum(L7)/10。その値をルート。



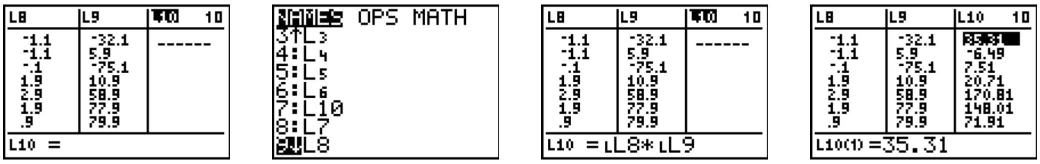
y　9　でL7を呼び出し
ます。

標準偏差は，60.50

3)　L8=L2−(L2 の平均)，L9=L3−(L3 の平均)。＜偏差X，偏差Yを求める＞

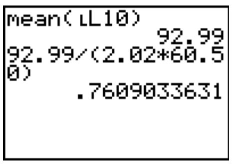


4)　L10=L8×L9　　＜偏差積を求める＞



5)　L10（偏差積）の平均を求める。

6)　相関係数＝(偏差積の平均)/{(X の標準偏差)×（Y の標準偏差）}



立川店の相関係数は 0.76
同様に，町田店の相関係数を求めると 0.96 となります。

±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

＜参考資料＞

統計学がわかる【回帰分析・因子分析編】　向後千春，富永敦子，技術評論社　2009