

年 組 番 氏名

SS 理数探究 I 統計処理② (分散、標準偏差)

○データの散らばり

5人ずつで作ったA、B、Cの3グループで、数学の試験をして次の得点を得たとします。

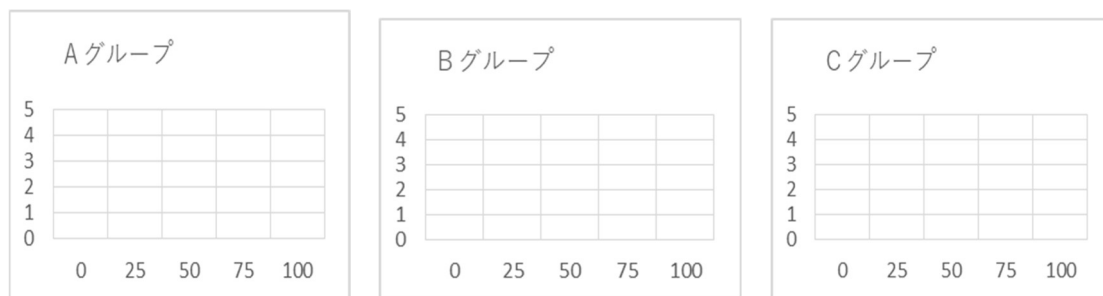
A : 25 50 50 50 75
 B : 50 50 50 50 50
 C : 0 0 50 100 100

3つのグループの平均点は A (), B (), C ()

このように、平均点は各グループの違いを教えるとは限りませんし、必ずしも平均点の前後に5人が集中しているわけではありません。

ではこのような各グループの点数の散らばりを目で見えてわかるようにするにはどうしたらいいでしょう。

グラフにするとよくわかります。A、B、Cのヒストグラム(度数分布図)をかいてみよう。



このヒストグラムより散らばりが一番小さいのは () グループ、散らばりが一番大きいのは () グループということがわかります。

では、散らばり度合いを数値で表現できないでしょうか。

その一つの方法として、データの各値と平均値との差を考えます。これを**偏差**といいます。

$$(\text{各値}) - (\text{平均値}) = (\text{偏差})$$

(各値が平均値より小さい場合は、偏差は負の数になります)

この偏差の平均値はどうなるでしょうか。

- A :
- B :
- C :

統計処理②

偏差の平均値はすべて（ ）なので、データの散らばり度合いを表すことはできません。
では、偏差の絶対値の平均値を求めてみましょう。

A :

B :

C :

このように計算した偏差の絶対値の平均値を**平均偏差**といいます。

この平均偏差は散らばりをよく表していますが、データが多くなると絶対値の計算が複雑になります。

そこで、偏差の2乗の平均値を考えてみましょう。

A :

B :

C :

このように計算した偏差の2乗の平均値を**分散**といい s^2 で表します。

分散もデータの散らばりをよく表しています。

しかし、単位を考えると、例えば測定単位が cm のとき、分散 s^2 の単位は cm^2 です。

そこで測定単位と同じ単位である分散の平方根 $\sqrt{s^2}$ を s で表し、データの**標準偏差**といいます。

このとき A、B、C の標準偏差はそれぞれ

$$A : \sqrt{250} = 5\sqrt{10} \doteq 15.81 \text{ (点)} \quad , \quad B : 0 \text{ (点)} \quad , \quad C : \sqrt{2000} = 20\sqrt{5} \doteq 44.72 \text{ (点)}$$

となり、平均点 50 点からの得点の散らばりが分散よりもイメージしやすくなります。

標準偏差の値が小さいほどデータの値は平均値の近くに集中し、逆に大きいほど平均値から離れたデータが多くなる傾向にあります。

統計処理②

○ファーストフード店のAバーガーとBバーガーのポテトは、太さがほぼ同じで本数も10本で同じです。しかし、Bバーガーのポテトの方が長いという噂があります。Aバーガーの店員はその噂が本当か確かめるため、2つの店のポテトの長さを測りました。

Aバーガーのポテト	
番号	長さ(cm)
1	4.1
2	6
3	4
4	4.3
5	5.1
6	6.1
7	5.2
8	5.3
9	4.4
10	5.5

Bバーガーのポテト	
番号	長さ(cm)
1	3.9
2	4.5
3	6.8
4	7.2
5	1.9
6	2.8
7	6.6
8	5.7
9	4.8
10	5.8

問1 噂を確かめるためには何を求めればよいですか。また、その噂が出た原因は何だと思いますか。

解答：

問2 エクセルで平均値と標準偏差を求めて、噂が正しいか調べてみよう。

(備考：エクセルで平均は=AVERAGE(:) 標準偏差は=STDEV.P(:)を用いる。)

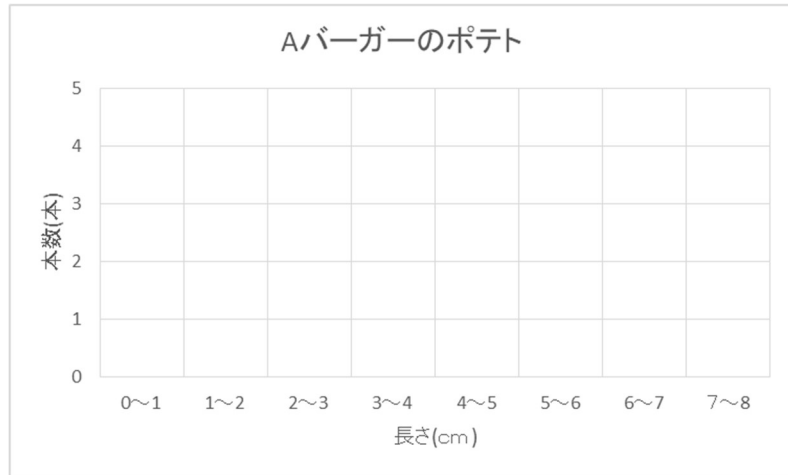
解答：

統計処理②

問3 AバーガーとBバーガーのポテトの長さの度数分布表とヒストグラムをかけ。

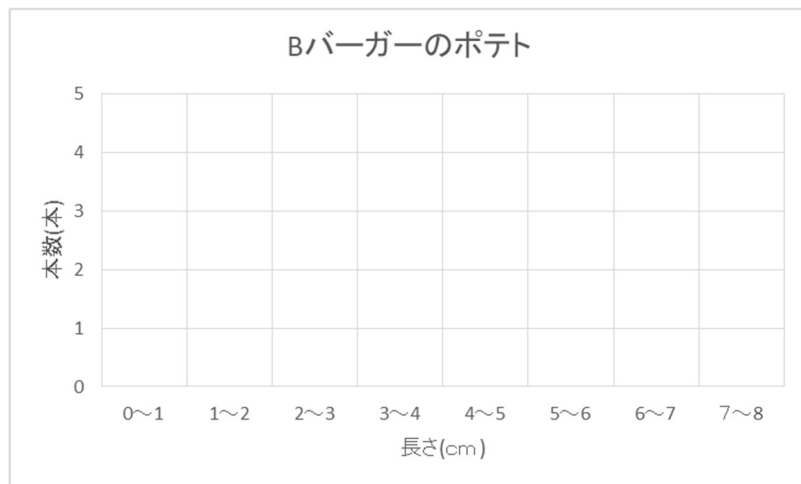
Aバーガーのポテト

階級〇以上～〇未満	
階級	度数
0～1	0
1～2	
2～3	
3～4	
4～5	
5～6	
6～7	
7～8	



Bバーガーのポテト

階級〇以上～〇未満	
階級	度数
0～1	0
1～2	
2～3	
3～4	
4～5	
5～6	
6～7	
7～8	



問4 標準偏差の値とヒストグラムの形は、問2と問3からどのような関係になっていますか。

解答：