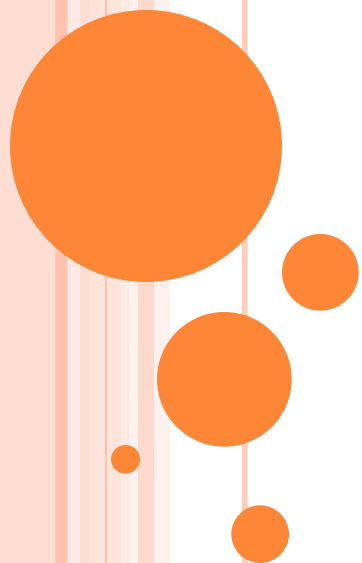


第1回

# 統計学

担当： 西山



# 今日の目標

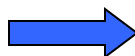
1. 統計的なものの見方とは？
2. 分布とは何か？
3. 分布の特徴とは何か？

# 最初のデータ解析

いくつかの数字が混ざっている様子を「**分布**」しています： 予習と復習の時間

**1、3、2、2、8、3、2、6、2、5**

データ全体から何が分かりますか？



最大は8で、最小は1である

2時間勉強する学生が多い

平均は3.4時間になる

個人差が大きく、ばらつきが目立つ



# 「データを読む」ということ

初歩的なデータ例： 予習と復習の予定時間

1、3、2、2、8、3、2、6、2、5



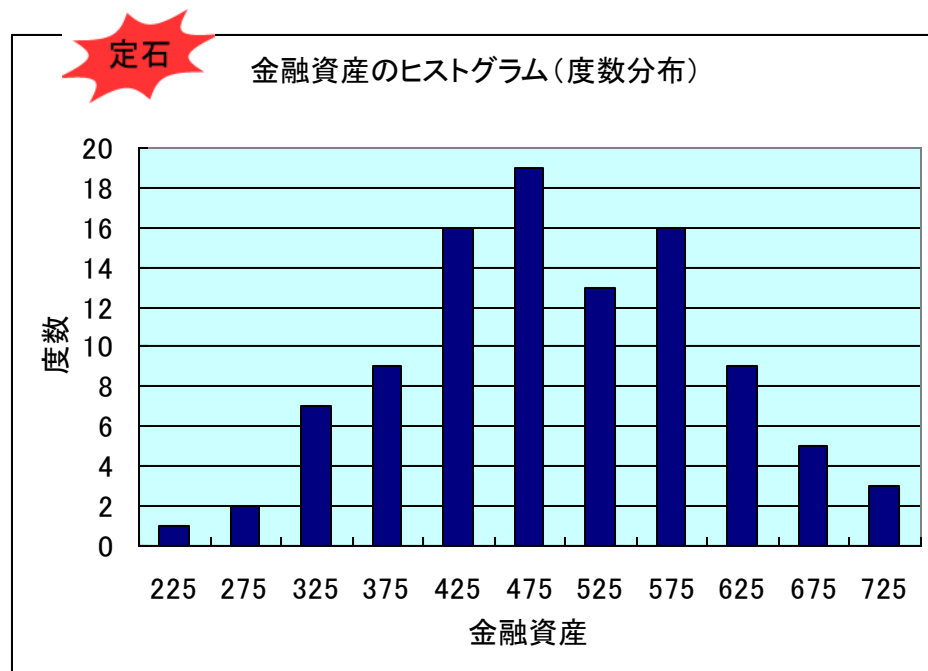
では次のデータはどうでしょう？

100世帯の金融資産保有額(単位:万円)

470	431	467	666	356	242	569	507	721	544
372	331	463	339	415	645	532	583	644	562
524	315	634	554	348	372	406	586	630	521
628	402	491	590	464	435	476	436	511	397
620	423	481	692	497	576	513	408	500	624
673	288	449	492	503	547	556	611	545	469
282	443	697	448	468	587	514	380	497	416
477	460	587	568	719	560	409	344	395	418
610	513	738	462	326	363	688	571	323	457
391	463	435	576	426	388	549	564	583	455

# 分布の特徴とは何か？

分布の特徴が分かれば、元のデータの形を想像できるはずです。



分布のグラフをなくしたら  
どうするか？

グラフの特徴だけを覚えて  
おけばよい

中心と広がりに着目

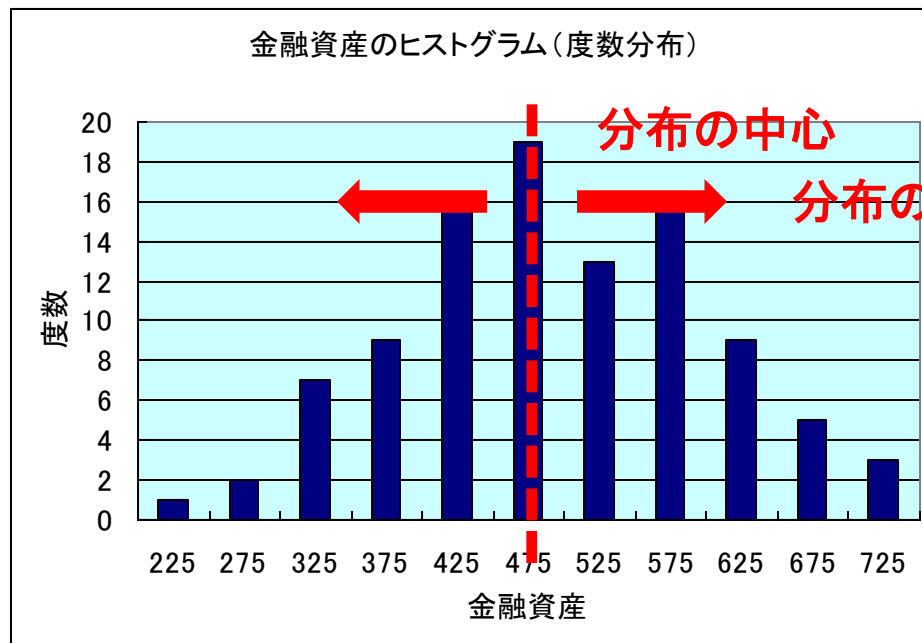
# 分布の特徴をつかむ定番

- 分布の中心の決め方

モード、メディアン、**平均値**

- 分布のばらつきの決め方

範囲、平均(絶対)偏差、**標準偏差**



**偏差**とはデータと平均  
値との差のこと！

## ばらつき(散らばり)の分析

	値	偏差	絶対偏差	二乗偏差
	1	-2.4	2.4	5.76
	3	-0.4	0.4	0.16
	2	-1.4	1.4	1.96
	2	-1.4	1.4	1.96
	8	4.6	4.6	21.16
	3	-0.4	0.4	0.16
	2	-1.4	1.4	1.96
	6	2.6	2.6	6.76
	2	-1.4	1.4	1.96
	5	1.6	1.6	2.56
合計	34	0	17.6	44.4
平均	3.4	0	1.76	4.44

標準偏差 = ルート分散  $\sqrt{4.44} = 2.11$

# 大きさとばらつきが統計の基本

## 元のデータ

クラスA

1  
3  
2  
2  
8  
3  
2  
6  
2  
5

3.4

クラスB

2  
2  
2  
2  
2  
2  
2  
2  
2  
2

2

クラスC

3.4  
3.4  
3.4  
3.4  
3.4  
3.4  
3.4  
3.4  
3.4  
3.4

3.4

クラスD

0  
0  
0  
0  
0  
6.8  
6.8  
6.8  
6.8  
6.8

3.4

データから  
平均を引くと

←平均値

偏差をそのまま  
平均しても常に  
ゼロです

## 偏差の状況(偏差=データ-平均値)

クラスA

-2.4  
-0.4  
-1.4  
-1.4  
4.6  
-0.4  
-1.4  
2.6  
-1.4  
1.6

クラスB

0  
0  
0  
0  
0  
0  
0  
0  
0  
0

クラスC

0  
0  
0  
0  
0  
0  
0  
0  
0  
0

クラスD

-3.4  
-3.4  
-3.4  
-3.4  
-3.4  
3.4  
3.4  
3.4  
3.4  
3.4



# ばらつきは標準偏差で測る

## 二乗偏差の状況

クラスA	クラスB	クラスC	クラスD
5.76	0	0	11.56
0.16	0	0	11.56
1.96	0	0	11.56
1.96	0	0	11.56
21.16	0	0	11.56
0.16	0	0	11.56
1.96	0	0	11.56
6.76	0	0	11.56
1.96	0	0	11.56
2.56	0	0	11.56
4.44	0	0	11.56
2.11	0	0	3.40

平均二乗  
偏差のこと

←分散

←標準偏差

平均値も標準偏差もデータ全体をまとめたものです！



# これも「分散」と呼んでいます

平均二乗偏差の定義どおりなら、

$$S^2 = \frac{1}{N} \sum_{i=1}^N (X_i - \bar{X})^2$$

第4章では、こちらを主に使います

不偏分散、と呼んでいます

$$\hat{\sigma}^2 = \frac{1}{N-1} \sum_{i=1}^N (X_i - \bar{X})^2$$



# 理解度チェッククイズ

- 3個のデータ

64、70、76

の平均値と標準偏差を暗算で求めなさい。

- 日本の成人男性の身長分布について、その平均値と標準偏差はどの程度の値が大雑把に答えなさい。



# 練習問題

- (0) 5個のデータ、1, 2, 3, 4, 5の標準偏差を求めなさい。
- (1) すべてのデータに一定の数値(たとえば10、-10)を加えた場合、平均値はどのように変化するか？
- (2) すべてのデータに一定の数値(たとえば2、0.5)をかけた場合、平均値はどうなるか？
- (3) すべてのデータに一定の数値(たとえば10、-10)を加えた場合、標準偏差はどのように変化するか？
- (4) すべてのデータに一定の数値(たとえば2、0.5)をかけた場合、標準偏差はどうなるか？

ここまでできたら次の問題

- (5) 偏差の合計は常にゼロである。
- (6) 分散 = 二乗の平均 - 平均の二乗

データの公式  
と呼んでいます



# ゲタの公式—平均値—

例題1  
12頁

元のデータ値を $X$ としたとき

$$Y = a + bX$$

のようにして値 $Y$ を定義する。このとき

$$\bar{Y} = a + b\bar{X}$$

統計学の試験の平均点が50点、標準偏差は10点だった。  
下駄をはかせるため全学生の得点を1割増しと、更に10点を足した。  
平均点はいくらになるか？

# ゲタの公式ーばらつきー

例題4  
18頁

元のデータ値をXとしたとき

$$Y = a + bX$$

のようにして値Yを定義する。このとき

$$S_y^2 = b^2 \times S_x^2 \Rightarrow \text{足した数は無関係なし}$$

**標準偏差は分散の正のルート！**

統計学の試験の平均点が50点、標準偏差は10点だった。  
下駄をはかせるため全学生の得点を1割増しと、更に10点を足した。  
分散はいくらになるか？ 標準偏差はいくらになるか？

# 合計の公式—平均値—

練習問題1  
13頁

合計の平均は平均の合計である。

$$Z = X + Y \text{ ならば、 } \bar{Z} = \bar{X} + \bar{Y}$$

式で示すのは非常に簡単です

ある定期試験で英語の平均点が70点、数学の平均点が50点だった。二科目の合計点の平均点は何点か？



# 合計の公式ーばらつき

合計の分散は分散の合計」になるとは限らない！

XとYが独立のときだけ

XとYが独立ならば、 $Z = X + Y$ のとき

$$S_z^2 = S_x^2 + S_y^2$$

合計の分散がどうなるか分からないが正解！

英語	数学	合計点
高い	高い	極めて高い
普通	普通	普通
低い	低い	極めて低い

分散上昇

英語	数学	合計点
高い	低い	普通
普通	普通	普通
低い	高い	普通

分散縮小





これも大事なポイント

例題2  
13頁

例題5  
19頁

**(5) 偏差の合計は常にゼロである。**

**(6) 分散 = 二乗の平均 - 平均の二乗**

3個のデータ1, 2, 2の分散を  
求めなさい。

