

平成25年度AIRIS

## 統計の基礎(2)

・データの代表値

平成25年8月22日

防衛医学研究センター 感染症疫学対策研究官

教授 加來浩器 (KAKU KOKI)

## 2つの群のデータ比較

**A群**

平均  
不偏分散  
標準偏差

**B群**

平均  
不偏分散  
標準偏差

A群とB群の代表値(平均、不偏分散、標準偏差)で検討する

### A中学校1年男子の2つのクラスの身長を比較する

組-番号	身長
1-1	137.0
1-2	144.0
1-3	149.0
1-4	136.0
1-5	142.0
1-6	156.0
1-7	150.0
1-8	138.0
1-9	146.0
1-10	157.0
1-11	160.0
1-12	153.0
1-13	149.0
1-14	142.0
1-15	147.0
1-16	135.0

組-番号	身長
2-1	144
2-2	143
2-3	139
2-4	166
2-5	169
2-6	144
2-7	147
2-8	138
2-9	176
2-10	133
2-11	170
2-12	137
2-13	146
2-14	140
2-15	122
2-16	177

The screenshot shows an Excel spreadsheet with two columns of data for Class 1 and Class 2. Each column contains 16 rows of student height data. Below the data, there are summary statistics for each class, including sum, average, minimum, maximum, variance, and standard deviation. The formulas for variance and standard deviation are explicitly shown as  $\text{不偏分散} = (\text{身長} - \text{平均})^2 \text{の合計} / n - 1$  and  $\text{標準偏差} = \text{不偏分散の平方根}$ .

項目	1年男子	2年男子
計 (SUM)		
平均 (AVERAGE)		
最小値 (MIN)		
最大値 (MAX)		
不偏分散 (VAR)	$\text{不偏分散} = (\text{身長} - \text{平均})^2 \text{の合計} / n - 1$	$\text{不偏分散} = (\text{身長} - \text{平均})^2 \text{の合計} / n - 1$
標準偏差 (STDEV)	$\text{標準偏差} = \text{不偏分散の平方根}$	$\text{標準偏差} = \text{不偏分散の平方根}$

問 1組と2組でのデータについて検討せよ。

### 各組のデータの代表値

	1組	2組
個数		
平均		
最小値		
最大値		
不偏分散		
標準偏差(SD)		

問 1組と2組でのデータについて検討せよ。

### 各組のデータの代表値

	1組	2組
個数	16	16
平均	146.3	149.4
最小値	135.0	122.0
最大値	160.0	177.0
不偏分散	60.1	278.0
標準偏差(SD)	7.8	16.7

### 数学、英語、国語の点数を比較

学生番号	数学	英語	国語
1	70	76	62
2	72	64	52
3	62	60	58
4	78	67	56
5	60	72	63
6	75	71	51
7	68	75	67
8	59	80	47
9	78	65	60
10	80	69	54
11	65	73	59
12	81	68	55
13	62	70	57

平均			
最小			
最大			
不偏分散			
標準偏差			

問 それぞれの成績の特徴について、教科間で比較しながら説明せよ。

The screenshot shows the Excel spreadsheet with the following data:

学生番号	数学	英語	国語
1	70	76	62
2	72	64	52
3	62	60	58
4	78	67	56
5	60	72	63
6	75	71	51
7	68	75	67
8	59	80	47
9	78	65	60
10	80	69	54
11	65	73	59
12	81	68	55
13	62	70	57
平均			
最小			
最大			
不偏分散			
標準偏差			

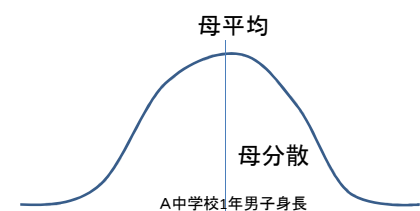
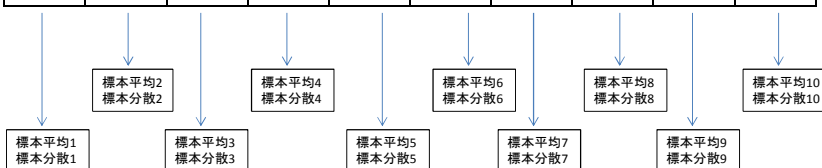
### 数学、英語、国語の点数を比較

学生番号	数学	英語	国語
1	70	76	62
2	72	64	52
3	62	60	58
4	78	67	56
5	60	72	63
6	75	71	51
7	68	75	67
8	59	80	47
9	78	65	60
10	80	69	54
11	65	73	59
12	81	68	55
13	62	70	57
平均	70	70	57
最小	59	60	47
最大	81	80	67
不偏分散	63.0	29.2	29.2
標準偏差	7.9	5.4	5.4

問 それぞれの成績の特徴について、教科間で比較しながら説明せよ。

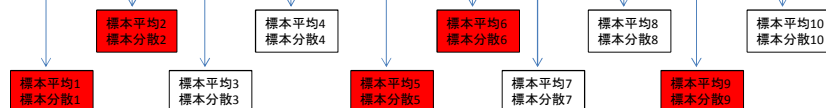
A中学校の1学年は10クラスある。  
いくつかのクラスのデータを抽出して1学年を推測

1組	2組	3組	4組	5組	6組	7組	8組	9組	10組
16人	16人	14人	18人	14人	16人	17人	15人	17人	15人



A中学校の1学年は10クラスある。  
いくつかのクラスのデータを抽出して1学年を推測

1組	2組	3組	4組	5組	6組	7組	8組	9組	10組
16人	16人	14人	18人	14人	16人	17人	15人	17人	15人

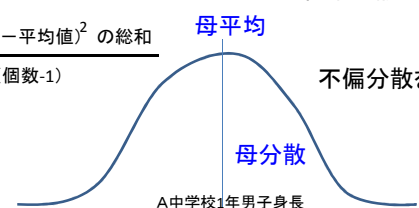


$$\text{標本分散} = \frac{(\text{データ}-\text{平均値})^2 \text{の総和}}{(\text{個数})}$$

$$\text{不偏分散} = \frac{(\text{データ}-\text{平均値})^2 \text{の総和}}{(\text{個数}-1)}$$

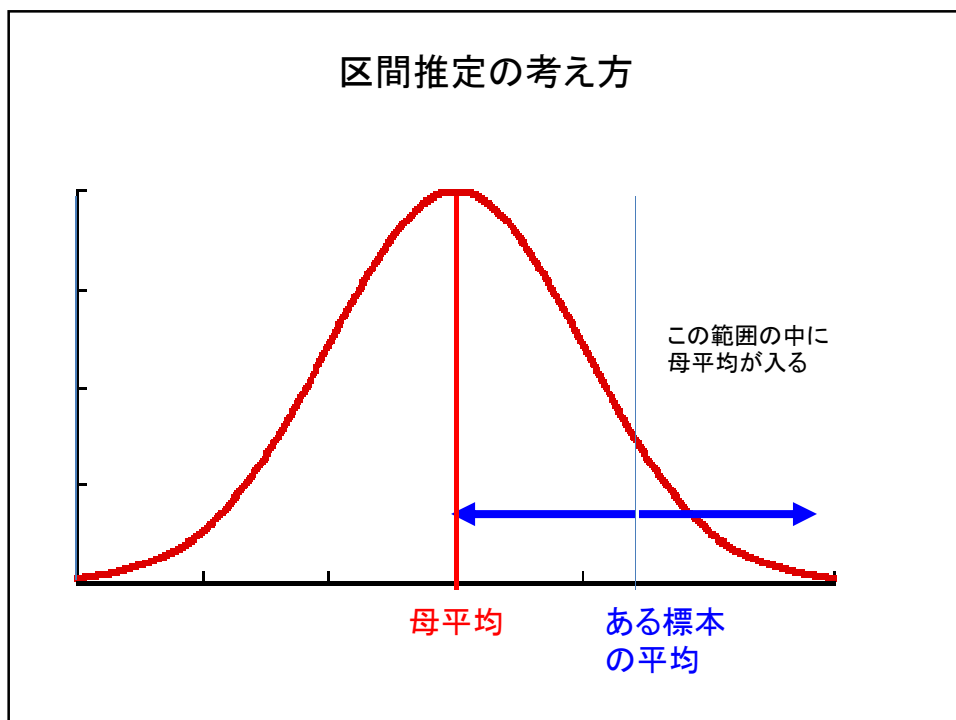
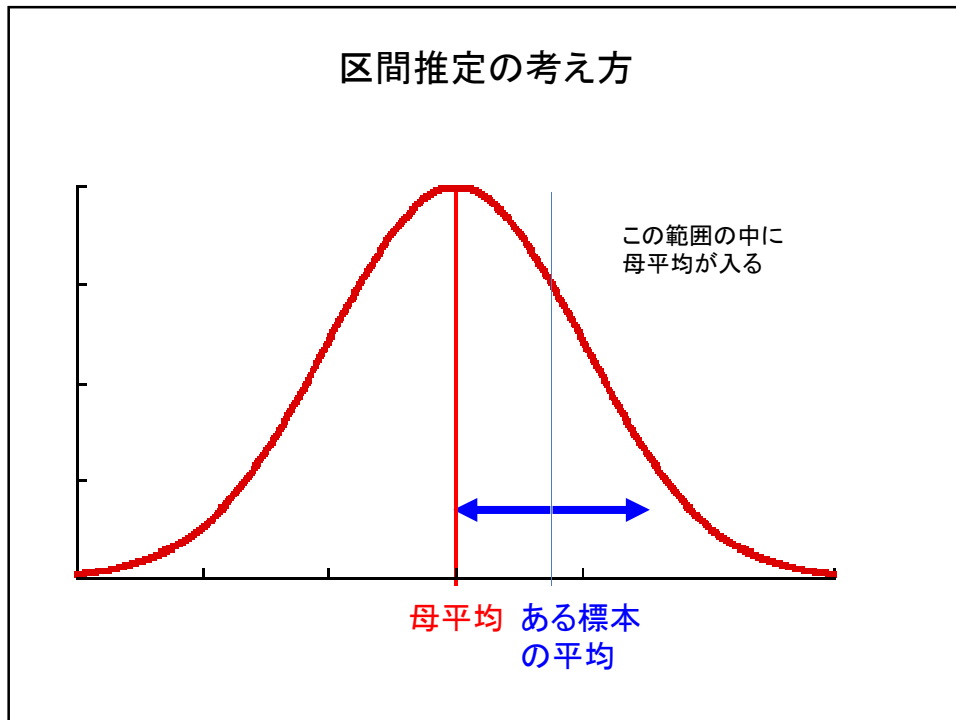
標本平均の平均は、母平均に一致する  
標本分散の平均は、母分散よりも少し小さい

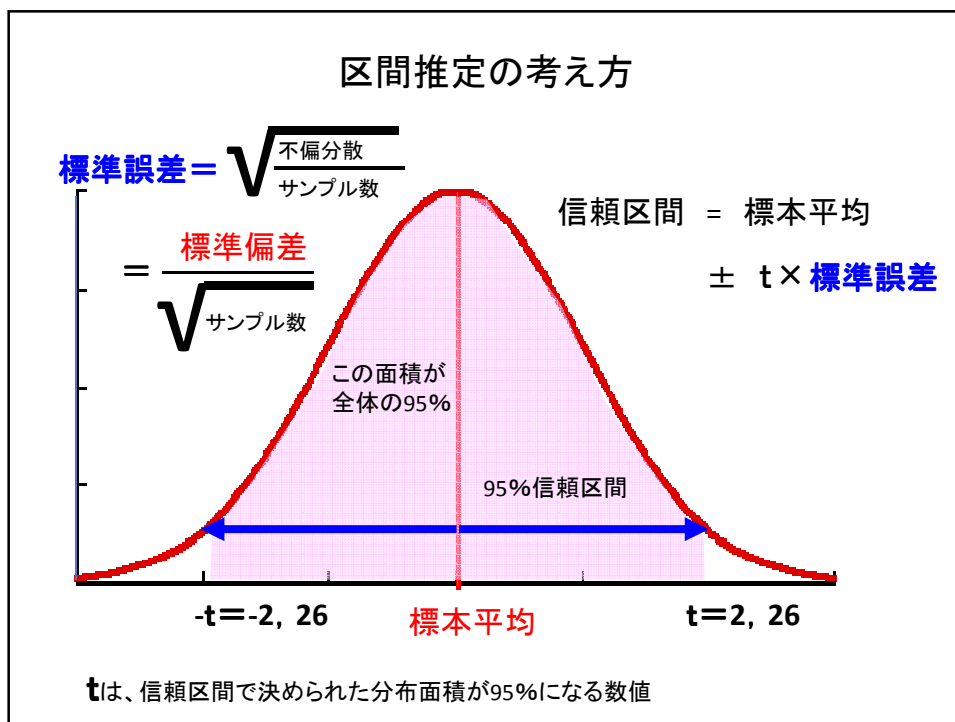
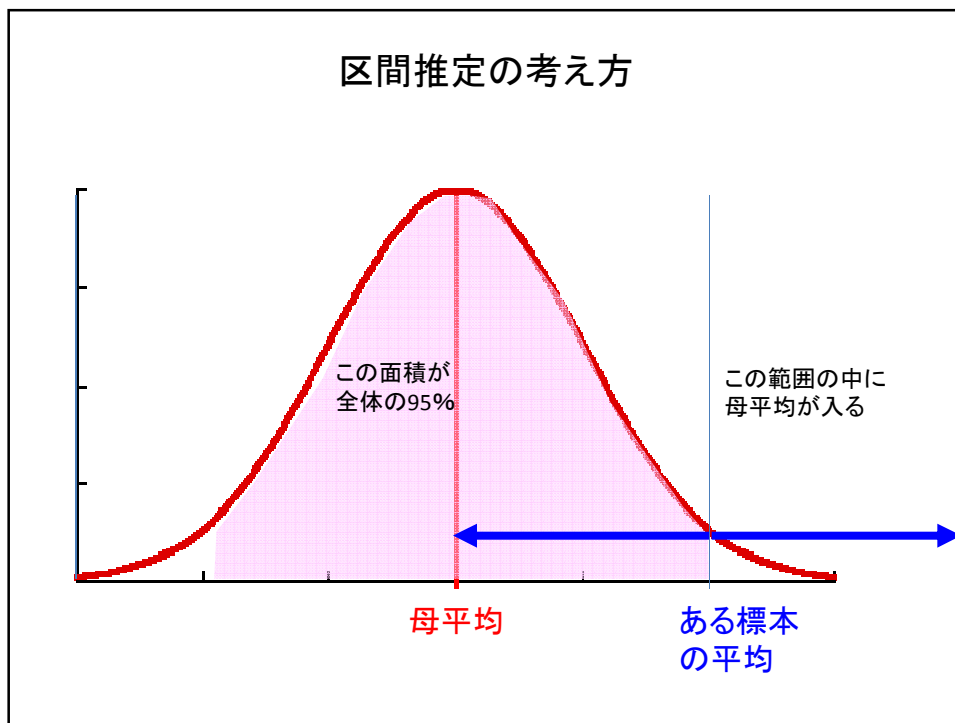
↓  
不偏分散を母分散の推定値として使用する



### 標本抽出(サンプリング)の考え方

- 母集団全体の平均(母平均)と分散(母分散)を知りたいときに、多くの場合は母集団全部のデータは手に入らない。
- そこで、母集団から無作為抽出をして標本のデータを手に入れます。
  - 標本の平均値は母平均の推定値として使える。
  - 標本の不偏分散は母分散の推定値として使える。







問3 以下のクラスを標本抽出した時の、A中学の1クラス当たりの男子生徒数を95%信頼区間で区間推定せよ。

1組	2組	3組	4組	5組	6組	7組	8組	9組	10組
16人	16人	14人	18人	14人	16人	17人	15人	17人	15人

1組、2組、5組、6組、9組をサンプル抽出した時

平均 =            不偏分散 =            標準誤差 =            95%CI =

1組、2組、3組、4組、5組をサンプル抽出した時

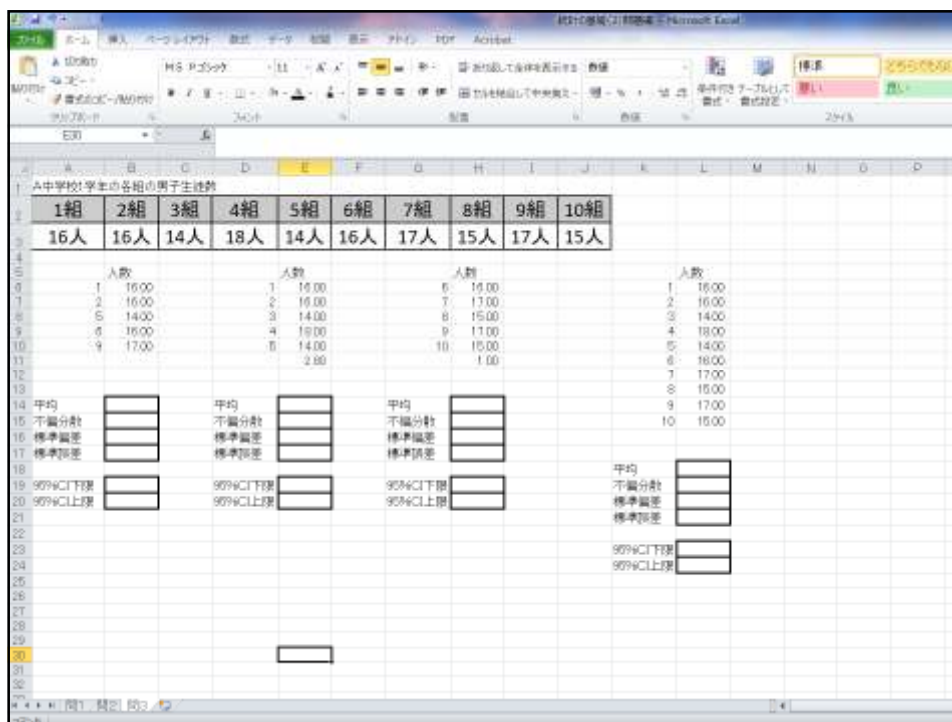
平均 =            不偏分散 =            標準誤差 =            95%CI =

6組、7組、8組、9組、10組をサンプル抽出した時

平均 =            不偏分散 =            標準誤差 =            95%CI =

全ての組を抽出した時

平均 =            不偏分散 =            標準誤差 =            95%CI =



問3 以下のクラスを標本抽出した時の、A中学の1クラス当たりの男子生徒数を95%信頼区間で区間推定せよ。

1組	2組	3組	4組	5組	6組	7組	8組	9組	10組
16人	16人	14人	18人	14人	16人	17人	15人	17人	15人

1組、2組、5組、6組、9組をサンプル抽出した時

平均 = 15.8    不偏分散 = 1.20    標準誤差 = 0.49    95%CI = 14.7~16.9

1組、2組、3組、4組、5組をサンプル抽出した時

平均 = 15.6    不偏分散 = 2.80    標準誤差 = 0.75    95%CI = 13.9~17.3

6組、7組、8組、9組、10組をサンプル抽出した時

平均 = 16.0    不偏分散 = 1.00    標準誤差 = 0.45    95%CI = 15.0~17.1

全ての組を抽出した時

平均 = 15.8    不偏分散 = 1.73    標準誤差 = 0.42    95%CI = 14.9~16.7