

---

# 講義9\_補足

## t検定について\_1

東京に住んでる10人の平均身長が170cmだった

大阪に住んでる10人の平均身長が179cmだった



大阪に住んでる人は身長が高そう

東京10人の身長
180
185
180
184
176
175
188
172
180
80

平均179

平均170

大阪10人の身長
243
171
172
172
174
175
168
169
170
176

... ?

大阪に住んでる人は  
本当に身長が高いのか？



t 検定

## t検定について\_2

東京10人の身長		大阪10人の身長
180	平均179	243
185		171
180		172
184		172
176		174
175	平均170	175
188		168
172		169
180		170
80		176

### t 検定

= 東京の平均と大阪の平均には  
差があるのか確認する検定



t検定するとp値という値が算出される。

(※)p値 = 東京の平均=大阪の平均の確率

ex)  $p=0.79$  : 79%の確率で平均は同じ

→ p値がすごく小さい(5%以下とか)  
= 平均が同じ確率は5%  
= 95%で東京と大阪の平均に差がある

(※)以降の説明は厳密には正しくない

# 前回課題のExercise.1について

0:コシヒカリの遺伝子, 1:ヒトメボレの遺伝子  
どの場所の遺伝子が葉の長さに影響を与えているのか知りたい。

個体No	葉の長さ	chr1_1	chr1_2	chr1_3	...
1	12.34	0	1	1	...
2	23.45	0	0	0	...
3	34.56	1	1	0	...
4	45.67	1	1	1	...
...					
197	98.76	1	1	1	...
198	87.65	1	0	0	...
199	76.54	1	0	0	...
200	65.43	0	0	1	...

葉の長さ	chr1_1
12.34	0
23.45	0
34.56	1
45.67	1
...	
98.76	1
87.65	1
76.54	1
15.43	0

# 前回課題のExercise.1について

0:コシヒカリの遺伝子, 1:ヒトメボレの遺伝子

葉の長さ	chr1_1
12.34	0
23.45	0
34.56	1
45.67	1
...	
98.76	1
87.65	1
76.54	1
15.43	0



chr1_1=0
12.34
23.45
15.43
11.43
...
11.11
22.22
33.33
15.43

平均  
87.65cm

平均  
45.67cm

chr1_1=1
34.56
45.67
98.76
76.54
...
99.99
88.88
77.77
66.66



t検定で  
差があるか確認

↓

差がある  
=  
遺伝子が変わる  
ことで葉の長さ  
に変化が

# 前回課題のExercise.1について

個体No	葉の長さ	chr1_1	chr1_2	chr1_3	...
1	12.34	0	1	1	...
2	23.45	0	0	0	...
3	34.56	1	1	0	...
4	45.67	1	1	1	...
...					
197	98.76	1	1	1	...
198	87.65	1	0	0	...
199	76.54	1	0	0	...
200	65.43	0	0	1	...

t検定 → 0/1で差がある？

t検定 → 0/1で差がある？

t検定 → 0/1で差がある？

差があるものを  
抜き出す

・  
・  
・

## 重回帰分析/前回課題のExercise.2について

遺伝子から肉量を説明したい。：重回帰分析

$$y = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 \dots + \beta_k x_k + e$$

牛No.	肉量	gene_A	gene_B	gene_C
1	100 kg	0	1	0
2	140 kg	1	1	0
3	200 kg	1	0	1
4	260 kg	0	1	1

0: gene\_Xを持っている  
1: gene\_Xを持っていない

$$y = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$$

例)

$$\text{肉量} = 40 \times \text{geneA} + 100 \times \text{geneB} + 160 \times \text{geneC}$$

## 重回帰分析/前回課題のExercise.2について

牛No.	肉量	gene_1	gene_2	gene_3	
1	100 kg	A	G	A	=40+50+10
2	140 kg	A	G	T	=40+50+50
3	200 kg	A	G	G	=40+50+110
4	260 kg	A	G	C	=40+50+170

$$y = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 \dots + \beta_k x_k + e$$

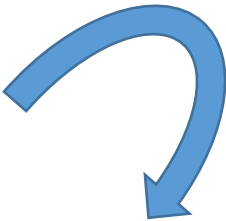
例) 肉量 =  $40(\text{gene}_1 = A\text{なら}) + 50(\text{gene}_2 = G\text{なら}) + 10(\text{gene}_3 = A\text{なら})$   
 $+ 50(\text{gene}_3 = T\text{なら}) + 110(\text{gene}_3 = G\text{なら}) + 170(\text{gene}_3 = C\text{なら})$

No. 1の個体なら、肉量=40+50+10=100



# 重回帰分析/前回課題のExercise.2について

牛No.	肉量	gene_1	gene_2	gene_3	
1	100 kg	A	G	A	=40+50+10
2	140 kg	A	G	T	=40+50+50
3	200 kg	A	G	G	=40+50+110
4	260 kg	A	G	C	=40+50+170



牛No.	肉量	gene_1A	gene_2G	gene_3A	gene_3T	gene_3G	gene_3C	
1	100 kg	1	1	1	0	0	0	=40+50+10
2	140 kg	1	1	0	1	0	0	=40+50+50
3	200 kg	1	1	0	0	1	0	=40+50+110
4	260 kg	1	1	0	0	0	1	=40+50+170

$y = \beta_1x_1 + \beta_2x_2 + \beta_3x_3 \dots + \beta_kx_k + e$

肉量 = 40×gene<sub>1A</sub> + 50×gene<sub>2G</sub> + 10×gene<sub>3A</sub> + 50×gene<sub>3T</sub> + 110×gene<sub>3G</sub> + 170×gene<sub>3C</sub>

例) No.2: 140 = 40\*1 + 50\*1 + 10\*0 + 50\*1 + 110\*0 + 170\*0

# 重回帰分析/前回課題のExercise.2について

牛No.	肉量	gene_1A	gene_2G	gene_3A	gene_3T	gene_3G	gene_3C	
1	100 kg	1	1	1	0	0	0	=40+50+10
2	140 kg	1	1	0	1	0	0	=40+50+40+10
3	200 kg	1	1	0	0	1	0	=40+50+100+10
4	260 kg	1	1	0	0	0	1	=40+50+160+10

前の式) 肉量 = 40×gene<sub>1A</sub> + 50×gene<sub>2G</sub> + 10×gene<sub>3A</sub> + 50×gene<sub>3T</sub> + 110×gene<sub>3G</sub> + 170×gene<sub>3C</sub>

$y = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 \dots + \beta_k x_k + e$

例) 肉量 = 40×gene<sub>1A</sub> + 50×gene<sub>2G</sub> + 40×gene<sub>3T</sub> + 100×gene<sub>3G</sub> + 160×gene<sub>3C</sub> + 10

# 重回帰分析/前回課題のExercise.2について

牛No.	肉量	gene_1A	gene_2G	gene_3T	gene_3G	gene_3C	...
1	100 kg	1	1	0	0	0	...
2	140 kg	1	1	1	0	0	...
3	200 kg	1	1	0	1	0	...
...	...	...	...	...	...	...	...
100	260 kg	1	1	0	0	1	...

$y = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 \dots + \beta_k x_k + e$

$X$

$100 = \beta_1 * 1 + \beta_2 * 1 + \beta_3 * 0 + \beta_4 * 0 + \beta_5 * 0 + \dots$   
 $140 = \beta_1 * 1 + \beta_2 * 1 + \beta_3 * 1 + \beta_4 * 0 + \beta_5 * 0 + \dots$   
 $200 = \beta_1 * 1 + \beta_2 * 1 + \beta_3 * 0 + \beta_4 * 1 + \beta_5 * 0 + \dots$   
...  
 $260 = \beta_1 * 1 + \beta_2 * 1 + \beta_3 * 0 + \beta_4 * 0 + \beta_5 * 1 + \dots$

になるような $\beta$ を求める。

# 重回帰分析/前回課題のExercise.2について

牛No.	肉量	gene_1A	gene_2G	gene_3T	gene_3G	gene_3C	...
1	100 kg	1	1	0	0	0	...
2	140 kg	1	1	1	0	0	...
3	200 kg	1	1	0	1	0	...
...	...	...	...	...	...	...	...
100	260 kg	1	1	0	0	1	...

出来るだけ  
差がないように  
 $\beta$ を求める

$y = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 \dots + \beta_k x_k + e$

$X$

$97 = \beta_1 * 1 + \beta_2 * 1 + \beta_3 * 0 + \beta_4 * 0 + \beta_5 * 0 + \dots$   
 $145 = \beta_1 * 1 + \beta_2 * 1 + \beta_3 * 1 + \beta_4 * 0 + \beta_5 * 0 + \dots$   
 $210 = \beta_1 * 1 + \beta_2 * 1 + \beta_3 * 0 + \beta_4 * 1 + \beta_5 * 0 + \dots$   
 $\dots$   
 $262 = \beta_1 * 1 + \beta_2 * 1 + \beta_3 * 0 + \beta_4 * 0 + \beta_5 * 1 + \dots$

実際には個体数より変数(x)の数の方が少ない場合、  
完璧な数式は出来ない。