

医療統計学

「医療統計学」

1 基本統計量 ヒストグラム

2 記述統計 推測統計

相関係数

3 仮説検定

× 二乗検定

t 検定

回帰分析

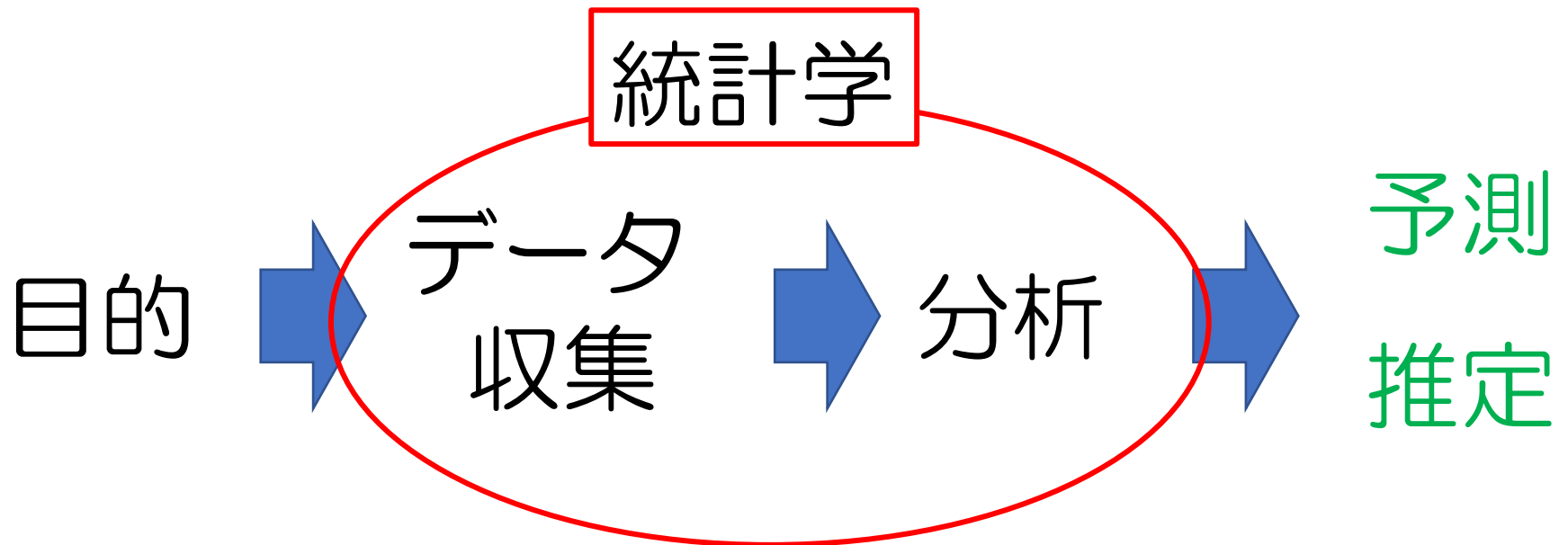


「統計学」とは

目的に応じてデータを収集して、

分析することで「予測」や「推定」

をする材料を作り出す学問



「医療統計学」

1 基本統計量 ヒストグラム

2 記述統計 推測統計

相関係数

3 仮説検定

× 二乗検定

t 検定

回帰分析



「統計」を進めて行く方法

目的に応じてデータを収集して分析

集めたデータ（数値）の特徴を

「パラメーター」や

「表・グラフ」

を使って分析する



あるテストの結果

54	48	90	78	81
63	70	62	77	83

これだけでは

データの特徴がわからないから

「数値」(特性値、統計量)

を出してみる

あるテストの結果

54	48	90	78	81
63	70	62	77	83

特性値(統計量)

最高 : 90 点

最低 : 48 点

平均 : 70.6点

中央値 : 73.5点

1 個 1 個のことを変量

「ちょうど真ん中の値」

2つあるときは

足して2で割る

「平均値」 $\bar{x}$

全部の変量を足して

サンプルの数で割ったもの

$$\bar{x} = \frac{x_0 + x_1 + x_2 + \sim x_n}{n}$$

$$\bar{x} = \frac{54 + 48 + 90 \sim + 83}{10} = 70.6$$

「中央値」 (メディアン)

変量を並べたとき

ちょうど真ん中の値

「2、5、8、10、11」なら「8」

「2、4、7、8、10、11」なら
「7.5」

「範囲」 (レンジ)

簡単に言うと「変量の幅」

変量の中の

「最大値」と「最小値」の引き算

「1、2、4、7、8、10、11」なら



「レンジ 10」

「度数分布表」

データを大まかな段階(階級)に分ける

点数（階級）		
0-19		
20-39		
40-59		
60-79		
80-100		



幅は自分で決めていい

「度数分布表」

それぞれの階級の真ん中の値を

点数（階級）	階級値	
0-19	10	
20-39	30	
40-59	50	
60-79	70	
80-100	90	



階級の真ん中

「度数分布表」

点数（階級）	階級値	度数
0-19	10	0
20-39	30	0
40-59	50	2
60-79	70	5
80-100	90	3

階級の中にある個数 ←

「度数分布表」

幅は自分で決めていい



点数（階級）	階級値	度数
0-19	10	0
20-39	30	0
40-59	50	2
60-79	70	5
80-100	90	3

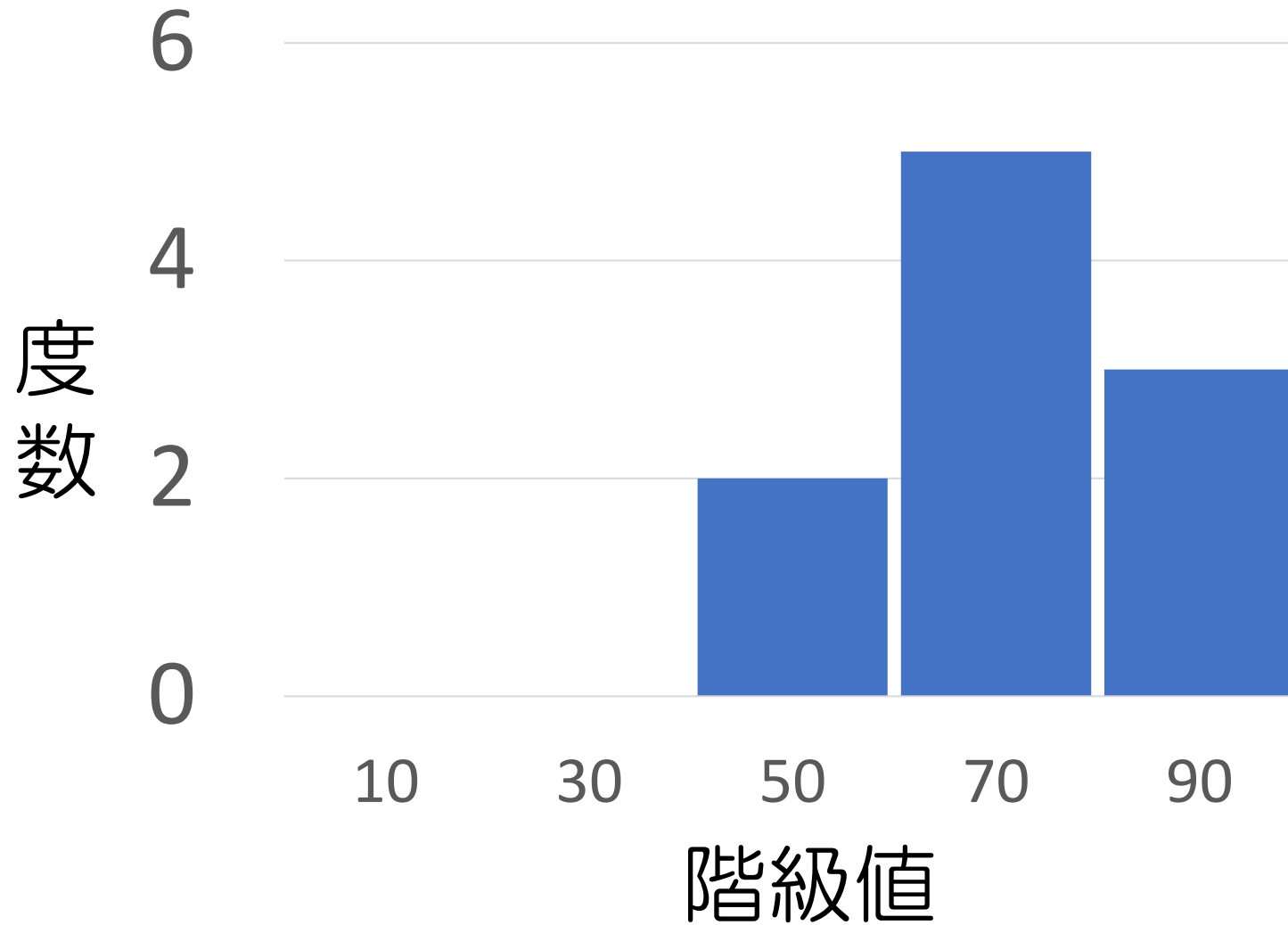


階級の真ん中



階級の中にある個数

「ヒストグラム」(棒グラフ)



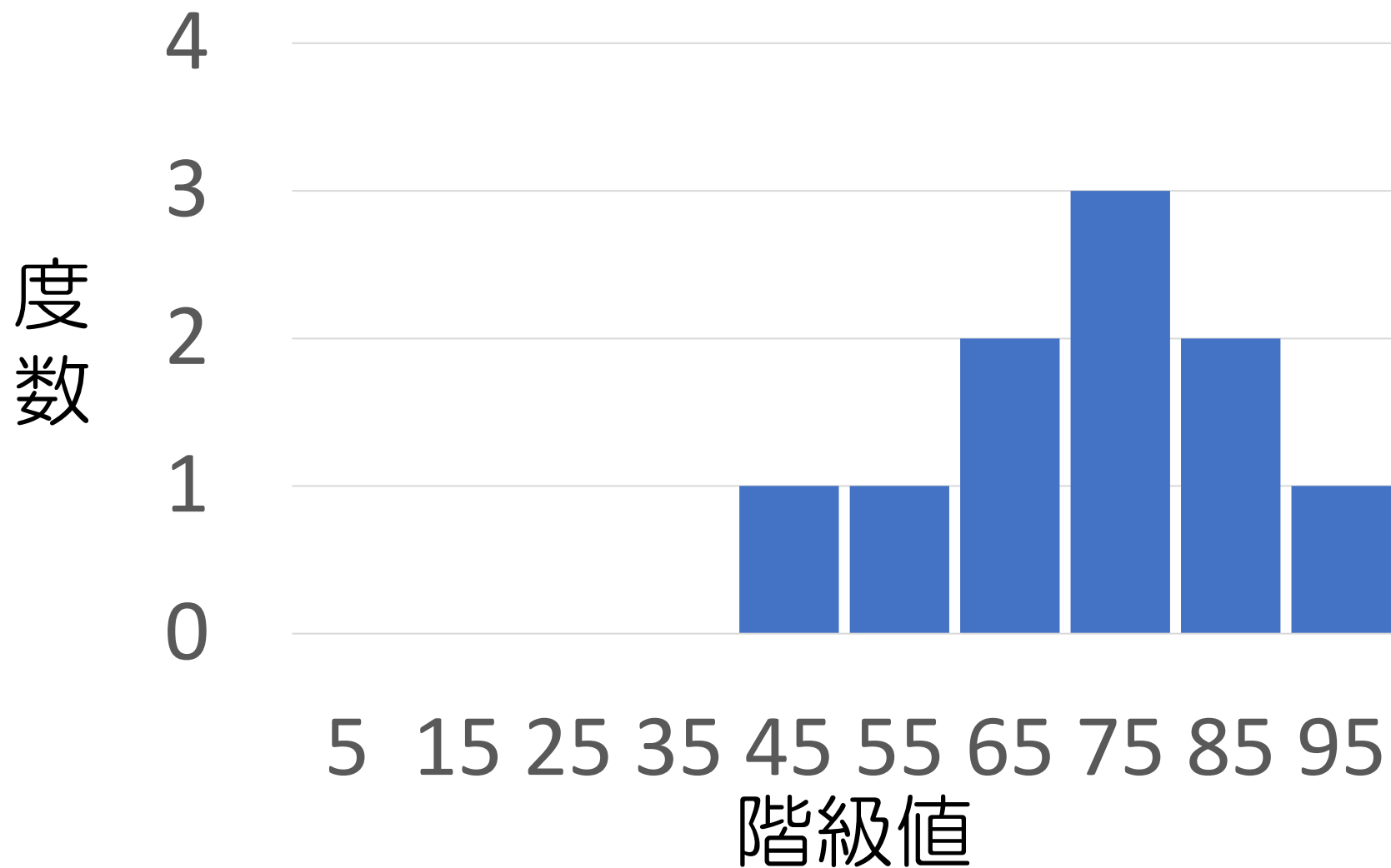
まだちょっとわかりにくい気がする！

「度数分布表」

階級を細かくすると

階級	階級値	度数
0-9	5	0
10-19	15	0
20-29	25	0
30-39	35	0
40-49	45	1
50-59	55	1
60-69	65	2
70-79	75	3
80-89	85	2
90-100	95	1

「ヒストグラム」(棒グラフ)

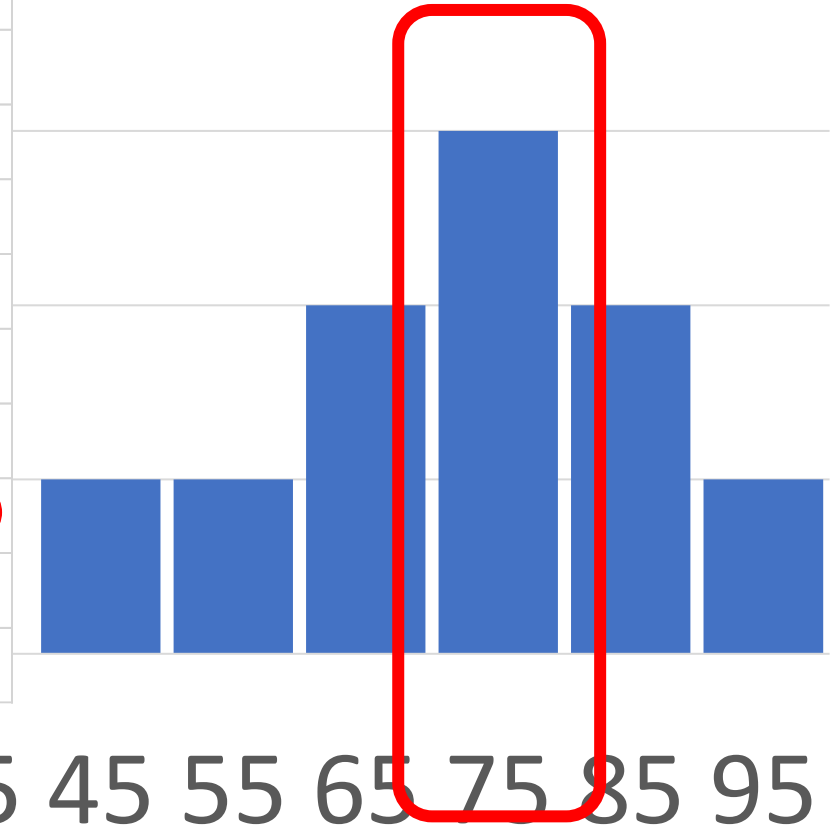


ちょっと細かくなった！

「最頻値」(モード)

階級	階級値	度数
0-9	5	0
10-19	15	0
20-29	25	0
30-39	35	0
40-49	45	1
50-59	55	1
60-69	65	2
70-79	75	3
80-89	85	2
90-100	95	1

5 15 25 35 45 55 65 75 85 95



1 番度数が大きいところの階級値

「度数分布表」

もう 1 歩進んで

階級	階級値	度数
0-9	5	0
10-19	15	0
20-29	25	0
30-39	35	0
40-49	45	1
50-59	55	1
60-69	65	2
70-79	75	3
80-89	85	2
90-100	95	1

「度数分布表」

全体を「1」としたときの割合

相対度数

階級	階級値	度数
0-9	5	0
10-19	15	0
20-29	25	0
30-39	35	0
40-49	45	1
50-59	55	1
60-69	65	2
70-79	75	3
80-89	85	2
90-100	95	1

0

0

0

0

0.1

0.1

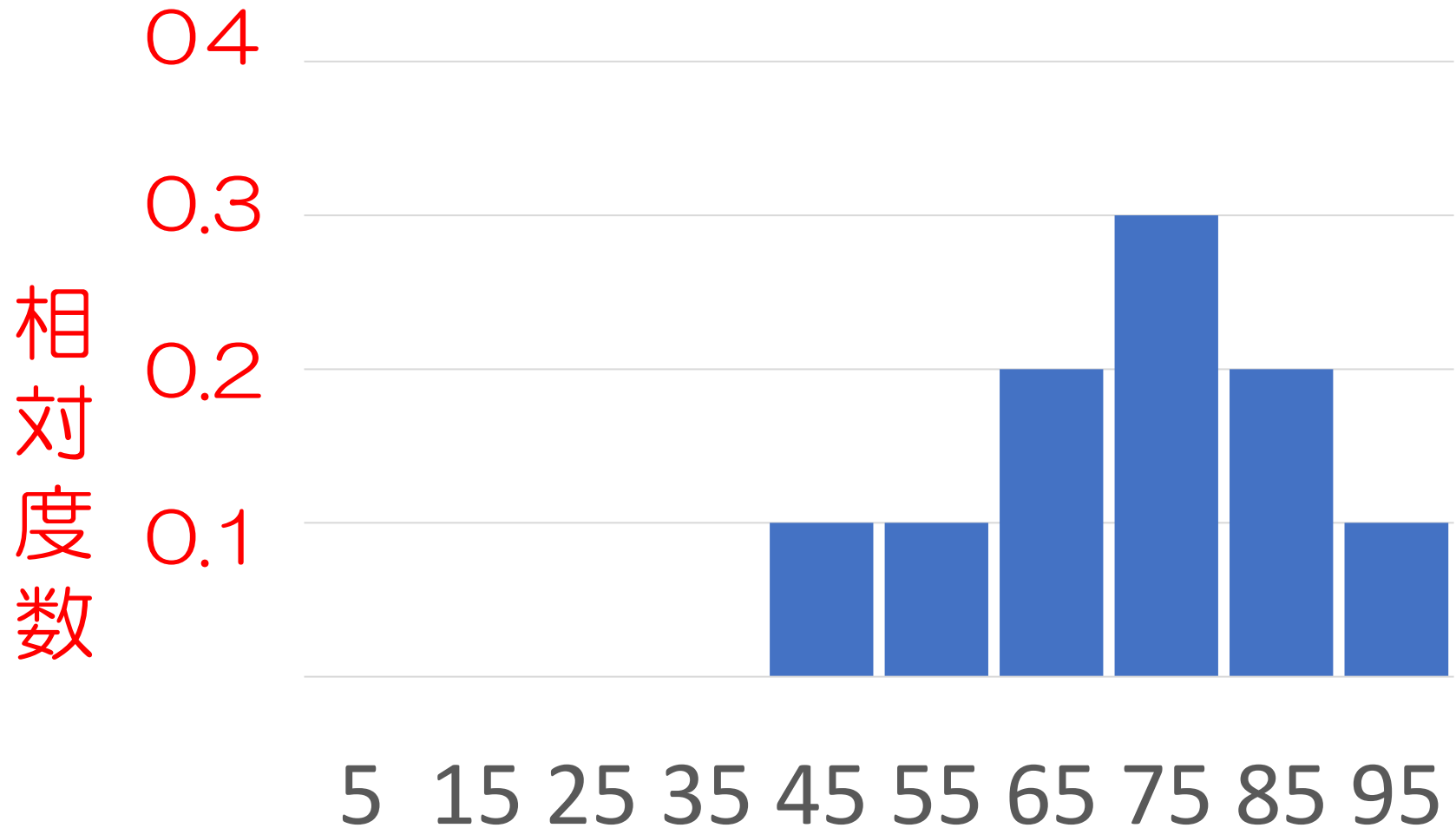
0.2

0.3

0.2

0.1

「ヒストグラム」(棒グラフ)



この棒グラフの面積を足すと「1」

「度数分布表」

今回は、10階級に分けたけど

どれくらいに分けたら

わかりやすい？

階級	階級値	度数
0-9	5	0
10-19	15	0
20-29	25	0
30-39	35	0
40-49	45	1
50-59	55	1
60-69	65	2
70-79	75	3
80-89	85	2
90-100	95	1



「階級の数」

大まかにやけど、

サンプル数が n 個の時

$\sqrt{n}$ 個の階級に分ければ
わかりやすい！

例えば、100個のサンプルなら「10階級」

「階級の数」

$\sqrt{n}$ 個の階級に分ける！

このとき、サンプルの中の

「最大値」から「最小値」を引いて

「階級数」で割ると

区切る目安になる！

今回なら

最高 : 90 点

最低 : 48 点

54	48	90	78	81
63	70	62	77	83

サンプル数10だから

$$\sqrt{10} \cong 3$$



$$\frac{90 - 48}{3} = 14$$

「14区切りで3個」

の階級にわければいい！

「度数分布表」

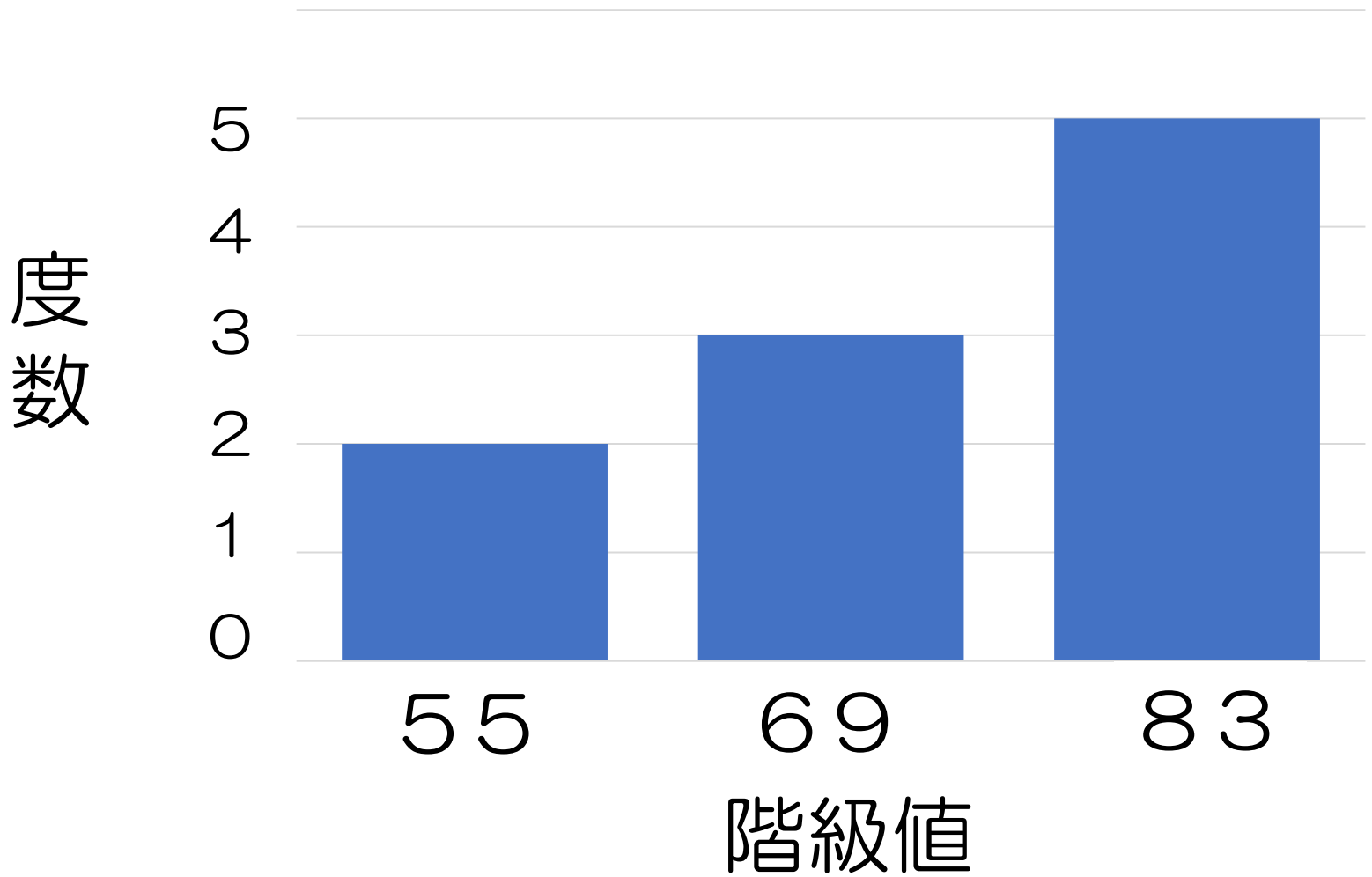
「1 4区切りで3個」

同じようにやって、

いるところだけとってみると

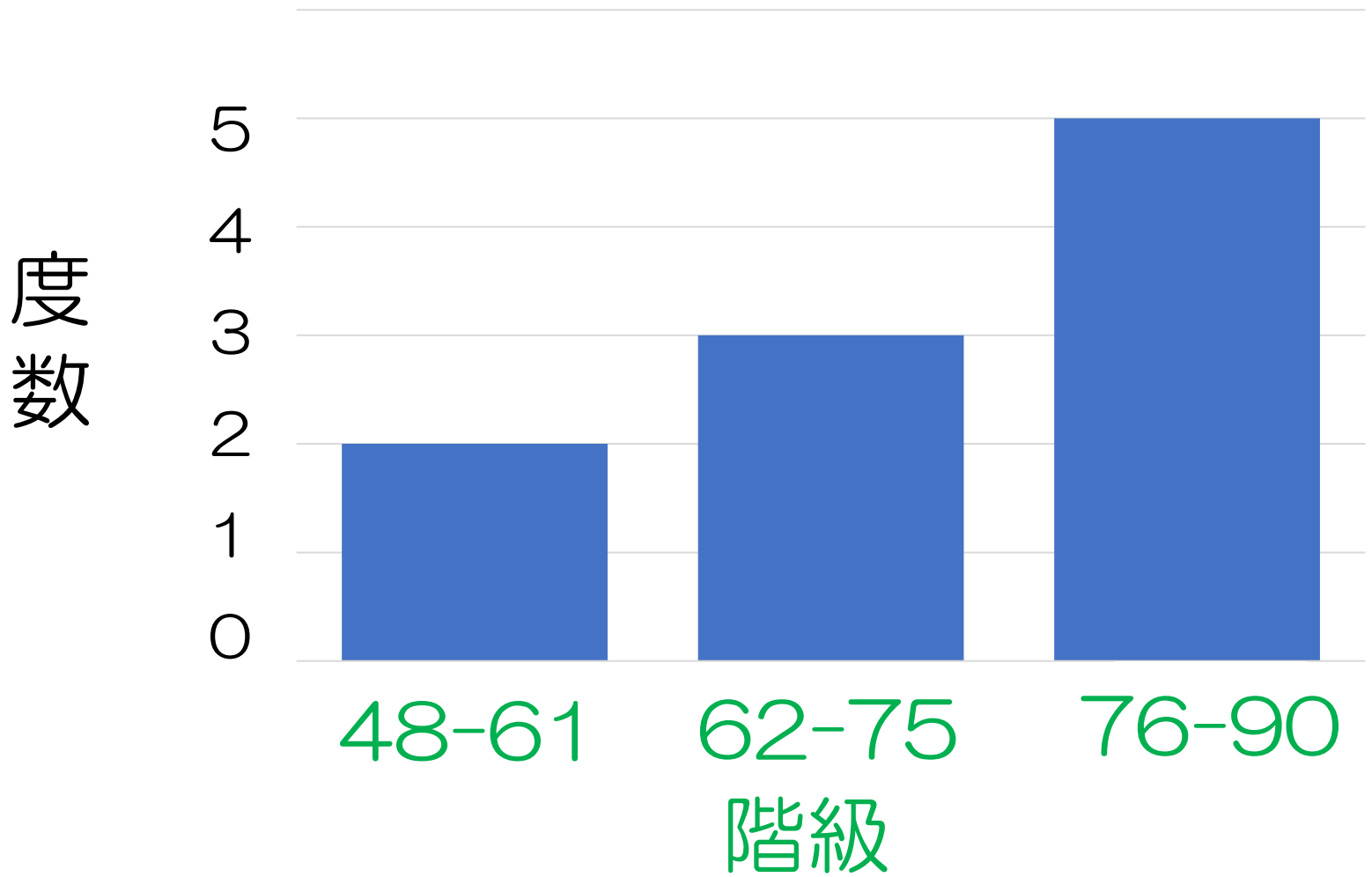
点数（階級）	階級値	度数
48-61	55	2
62-75	69	3
76-90	83	5

「ヒストグラム」(棒グラフ)



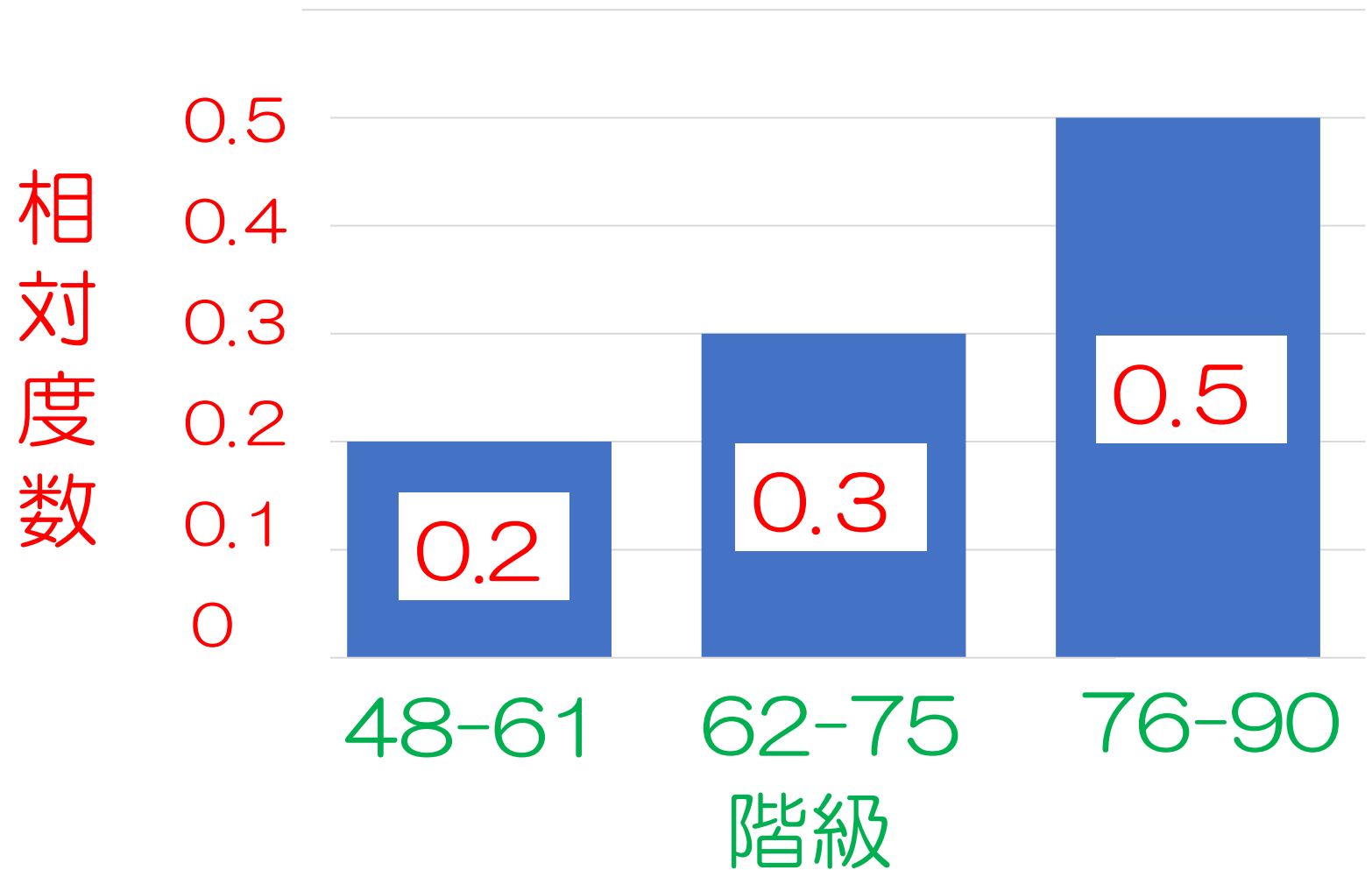
微妙に~~わかりにくい~~

「ヒストグラム」(棒グラフ)



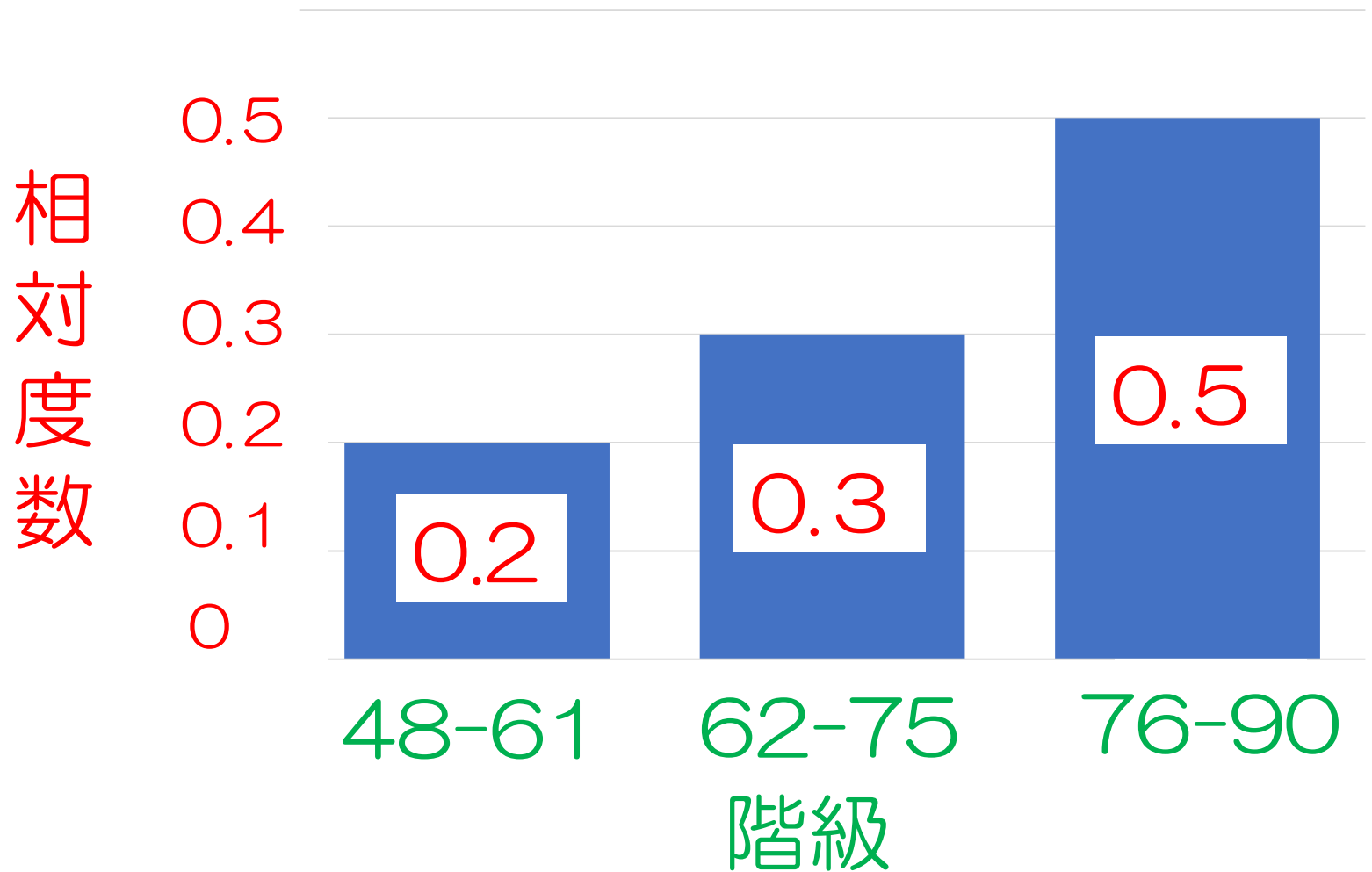
階級値を階級に変えてみる

「ヒストグラム」(棒グラフ)



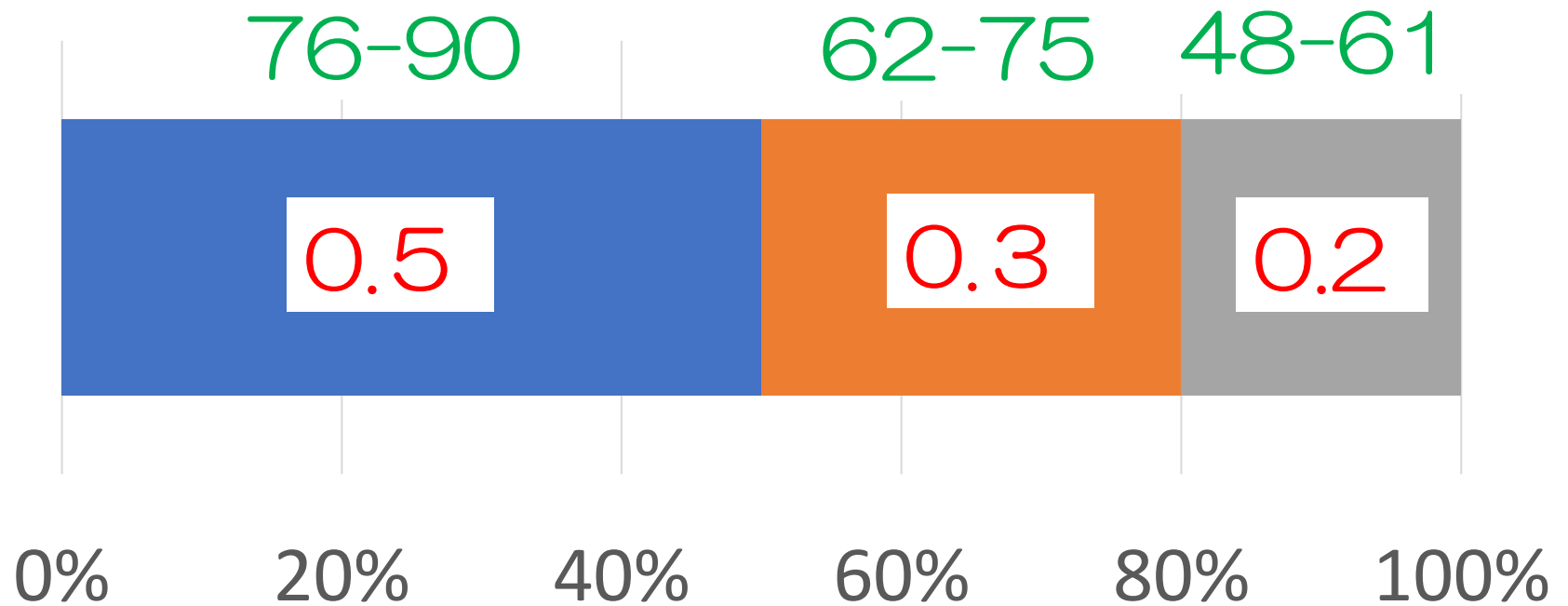
相対度数に変えてみる

「ヒストグラム」(棒グラフ)



面積が「1」になるんだから

「ヒストグラム」(帯グラフ)



こういうグラフもあり！

言いたいことが、一番わかるグラフを
選べばいい

グラフは、

「棒グラフ」

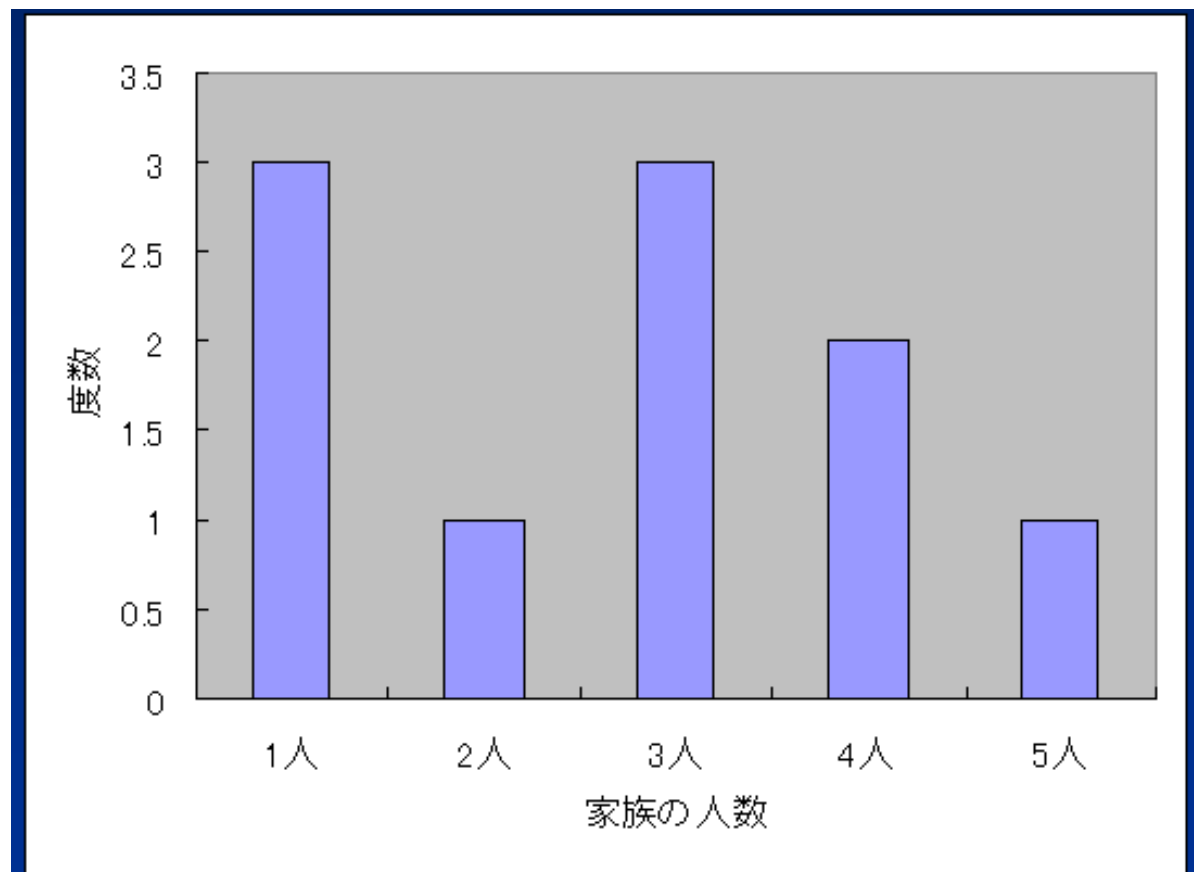
じゃないとあかんの？



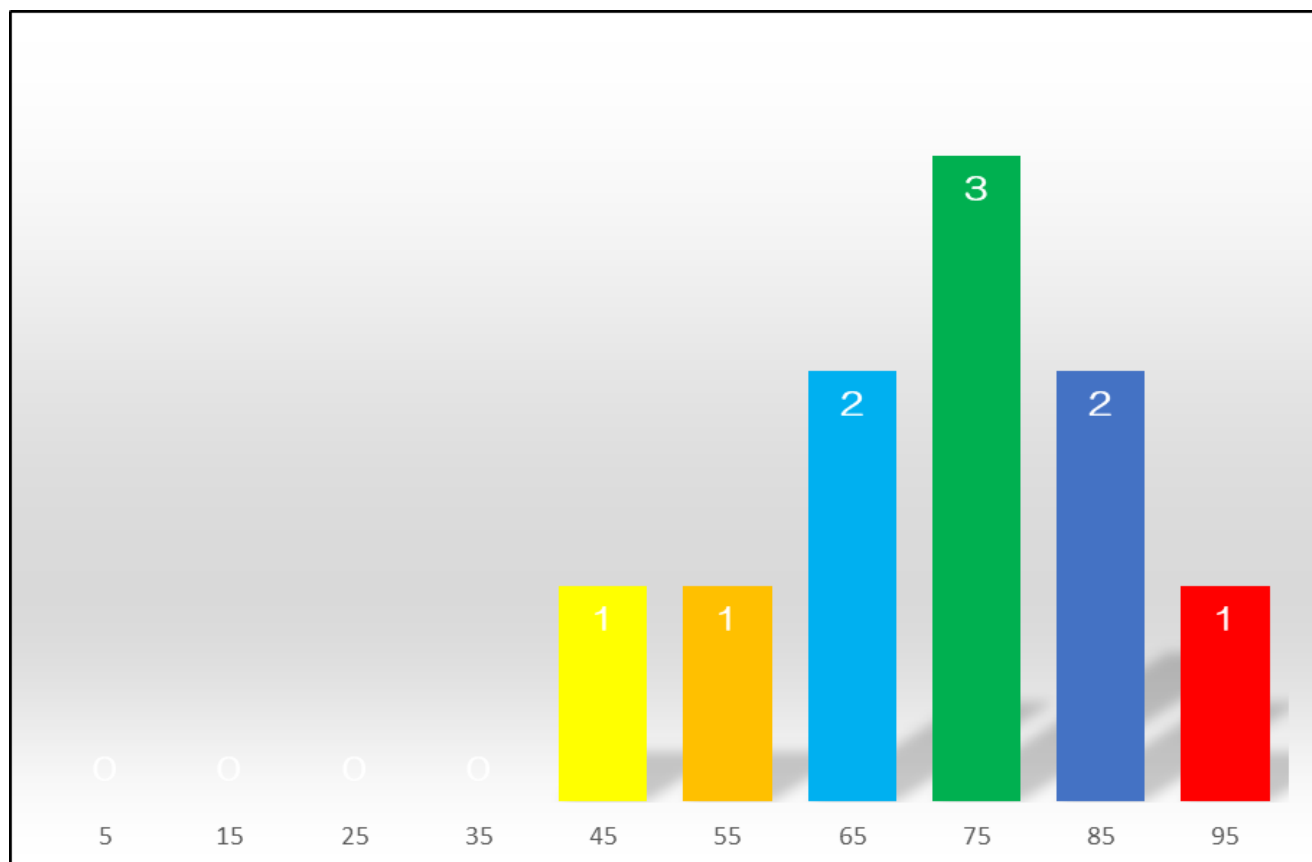
グラフは、何でもいい！

目的によってわかりやすい

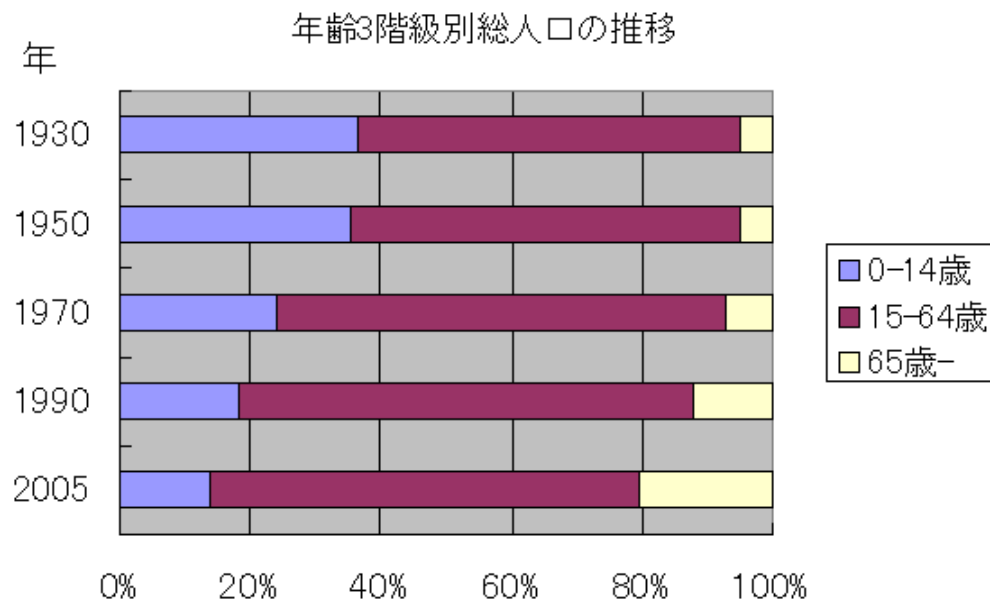
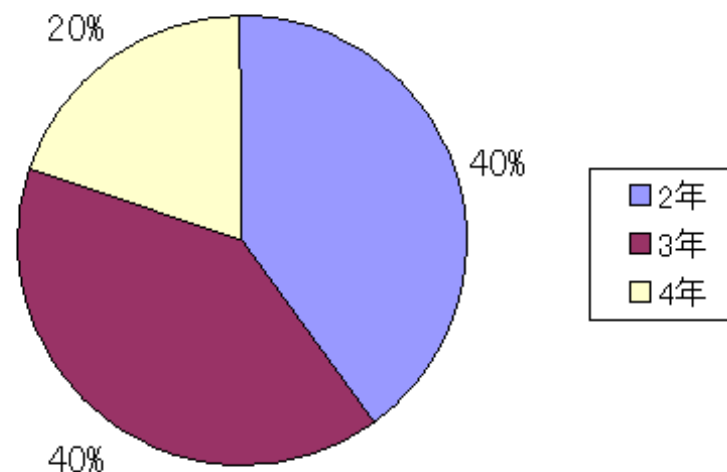
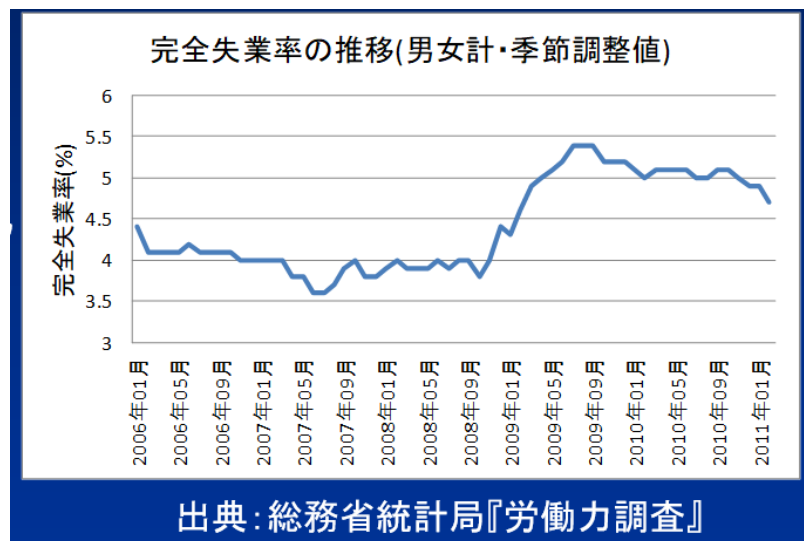
ものにしたらいいい！



こんな風に、
いろいろ変えまくっても

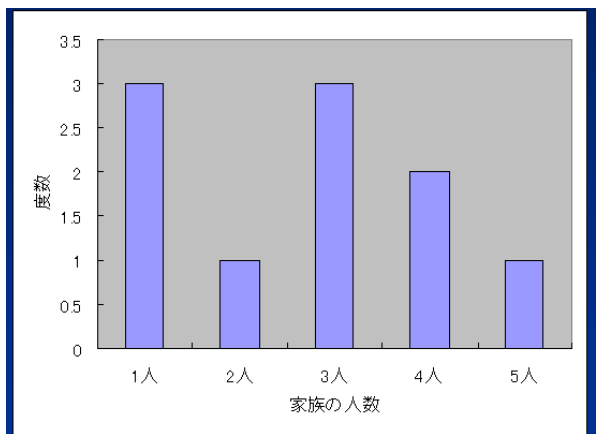


わかりやすければ何でもいい！

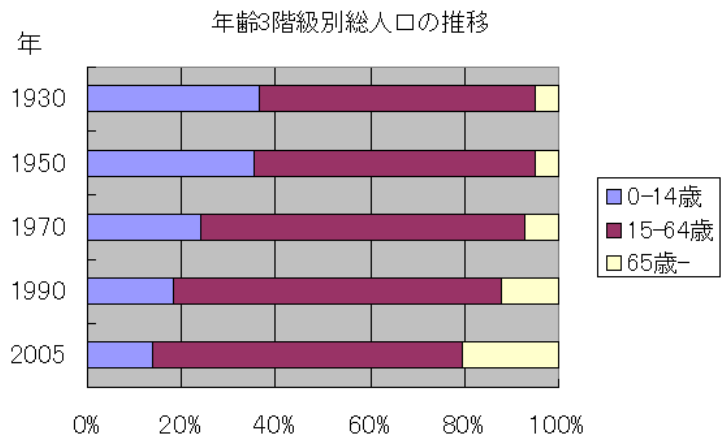


ある程度有名なものとしては

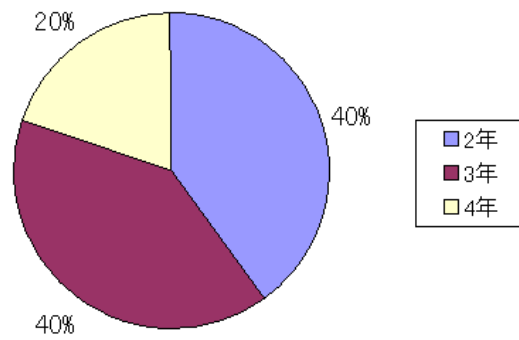
大小を比較したい
「棒グラフ」



%を比較したい
「帯グラフ」

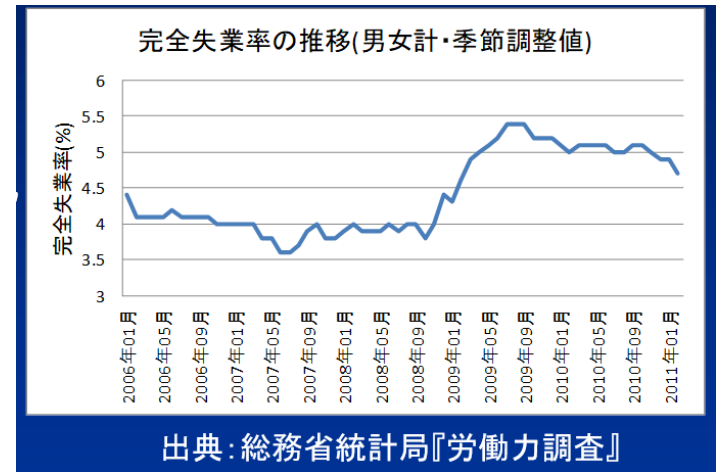


割合をみせたい
「円グラフ」

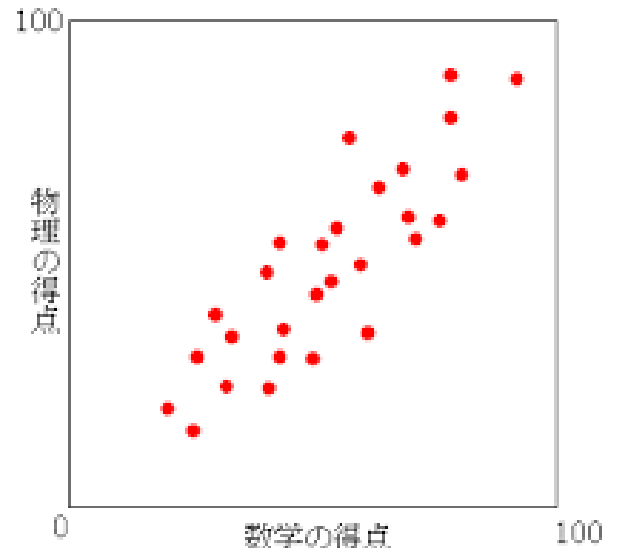


ある程度有名なものとしては

時間変化をみせたい
「折れ線グラフ」



2つの量の関係を見せたい
「散布図」



階級	階級値	度数
0-9	5	0
10-19	15	0
20-29	25	0
30-39	35	0
40-49	45	1
50-59	55	1
60-69	65	2
70-79	75	3
80-89	85	2
90-100	95	1

ところで

「度数分布表」だけで

分析を終わっていいの？

「平均」とか何にも
使っていないけど…



テスト1	20	20	0	0	0
------	----	----	---	---	---

テスト2	8	8	8	8	8
------	---	---	---	---	---

テスト3	18	13	6	2	1
------	----	----	---	---	---

どのテストも平均は「8点」

分析結果が、同じはずがない💡

他の分析方法が必要！



この点数を例に考えてみよう！

平均値「8」

2	5	9	11	13
---	---	---	----	----



平均より下だから
成績は悪い？



平均よりかなり上
成績はいい？

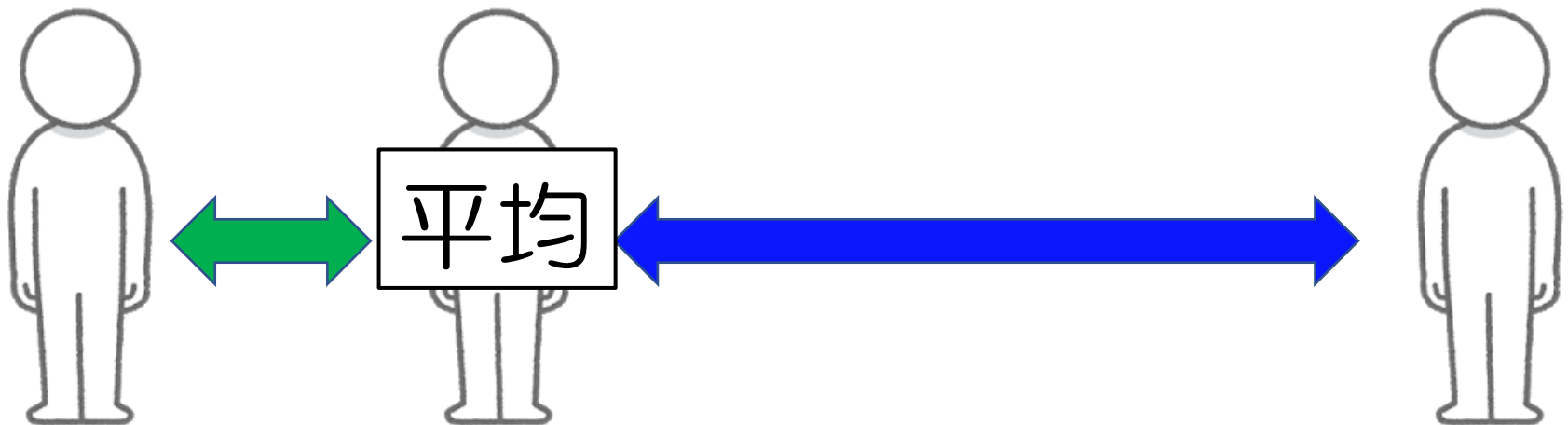
仮説

「偏差」

まずは、それぞれの変量が

平均値から

どれくらい離れてるのか



「偏差」

それぞれの変量の

平均値からの差

「2、5、9、11、13」

平均値「8」を引く ↓

「-6、-3、1、3、5」

「偏差」

「分散」

偏差 「-6、-3、1、3、5」

「マイナス」の値もあるから

比較しにくいなあ



「分散」

偏差 「-6、 -3、 1、 3、 5」

↑ ↑

「-」があるからわかりにくい！

全部 「2乗」にして「プラス」に
してしまう！！



「分散」

偏差 「-6、-3、1、3、5」



2乗する

「36、9、1、9、25」

これで分かりやすくなった！



「分散」

「36、 9、 1、 9、 25」

これ全部足してサンプル数で割ると…

全体のばらつき具合がわかる！

$$\frac{36 + 9 + 1 + 9 + 25}{5} = 16$$

(ばらつき具合の平均)

「分散」

全体のばらつき具合の平均

$$\frac{36 + 9 + 1 + 9 + 25}{5} = 16$$

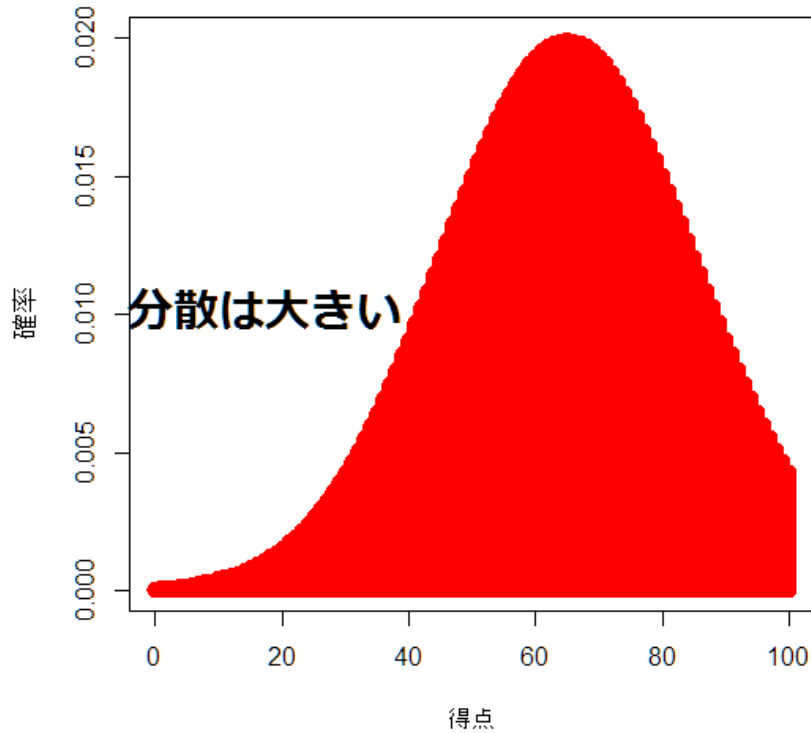
これを「分散」

分散が大きい $\Rightarrow$ データがばらついてる

分散が小さい $\Rightarrow$ 〃 ばらついてない

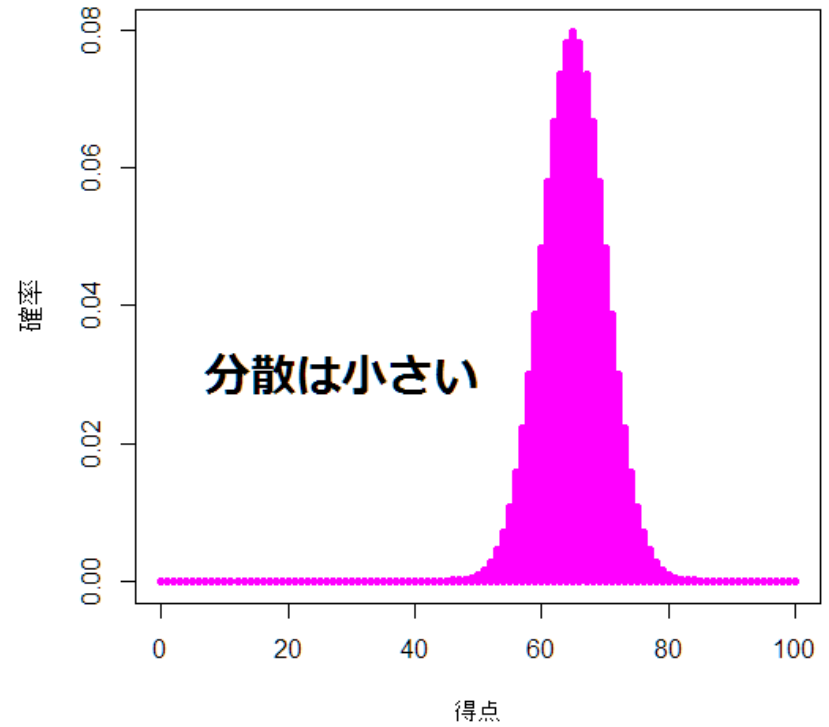
「分散」

分散が**大**



ばらついてる

分散が**小**



ばらついてない

「分散」

「分散が16」はわかったけど…

あんまり意味がわからん！

ばらつきが大きい？ 小さい？

そもそも単位は何？



わかりにくいのは
2乗したせい！

「標準偏差」

2乗したものを元に戻してみよう

「分散 16」

$$\sqrt{16} = 4$$

これを「標準偏差」

単位が「点数」に

点数のばらつき具合がわかった！

結局何がわかったのか

2	5	9	11	13
---	---	---	----	----

平均値「8」

$$\sqrt{16} = 4$$

「標準偏差」

$$16$$

「分散」

結局何がわかったのか

2	5	9	11	13
---	---	---	----	----

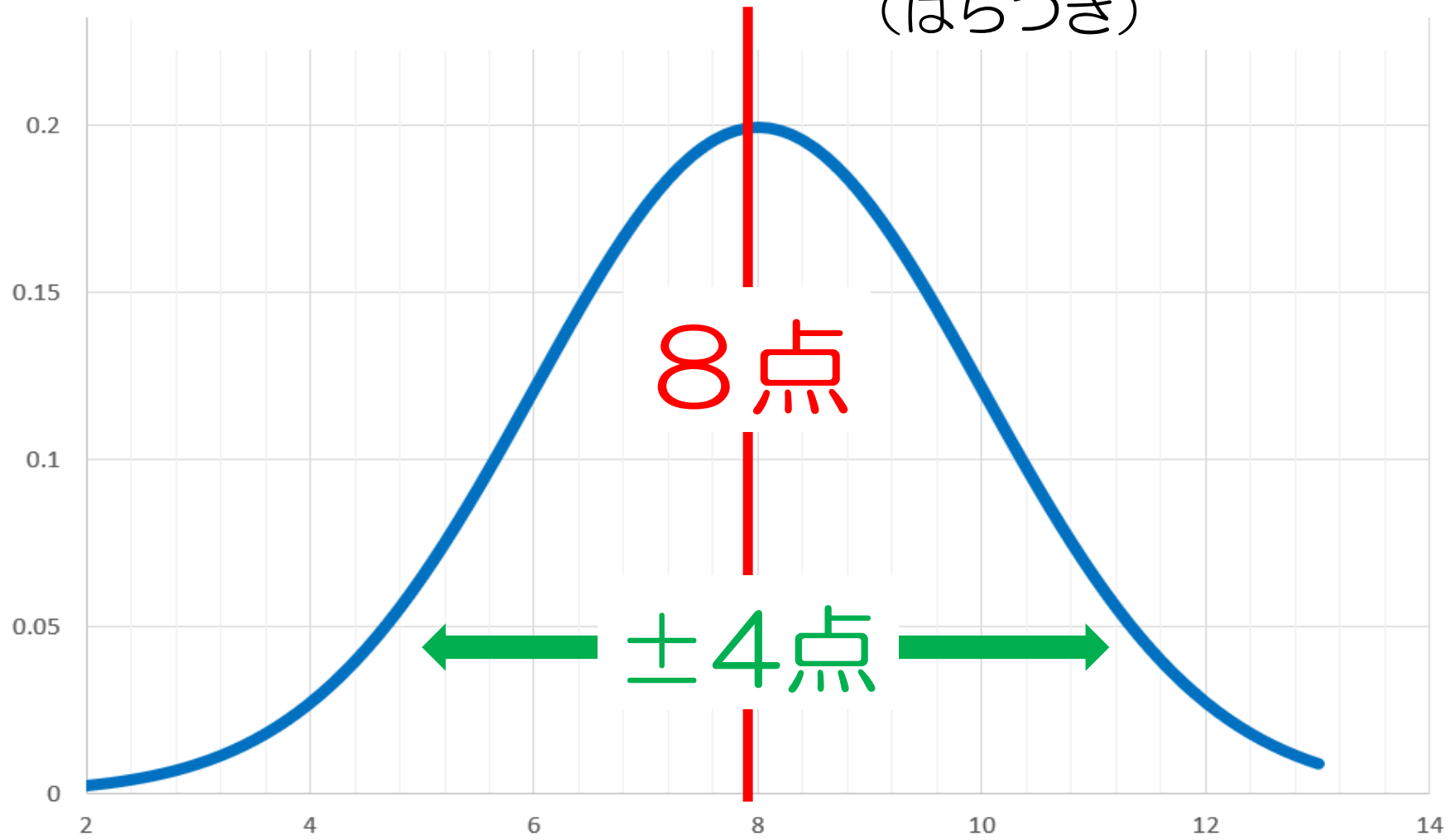
「平均値 8点、標準偏差 4点」

のテストという事がわかった！

グラフでみてみると

「平均値 **8点**、標準偏差 **4点**」

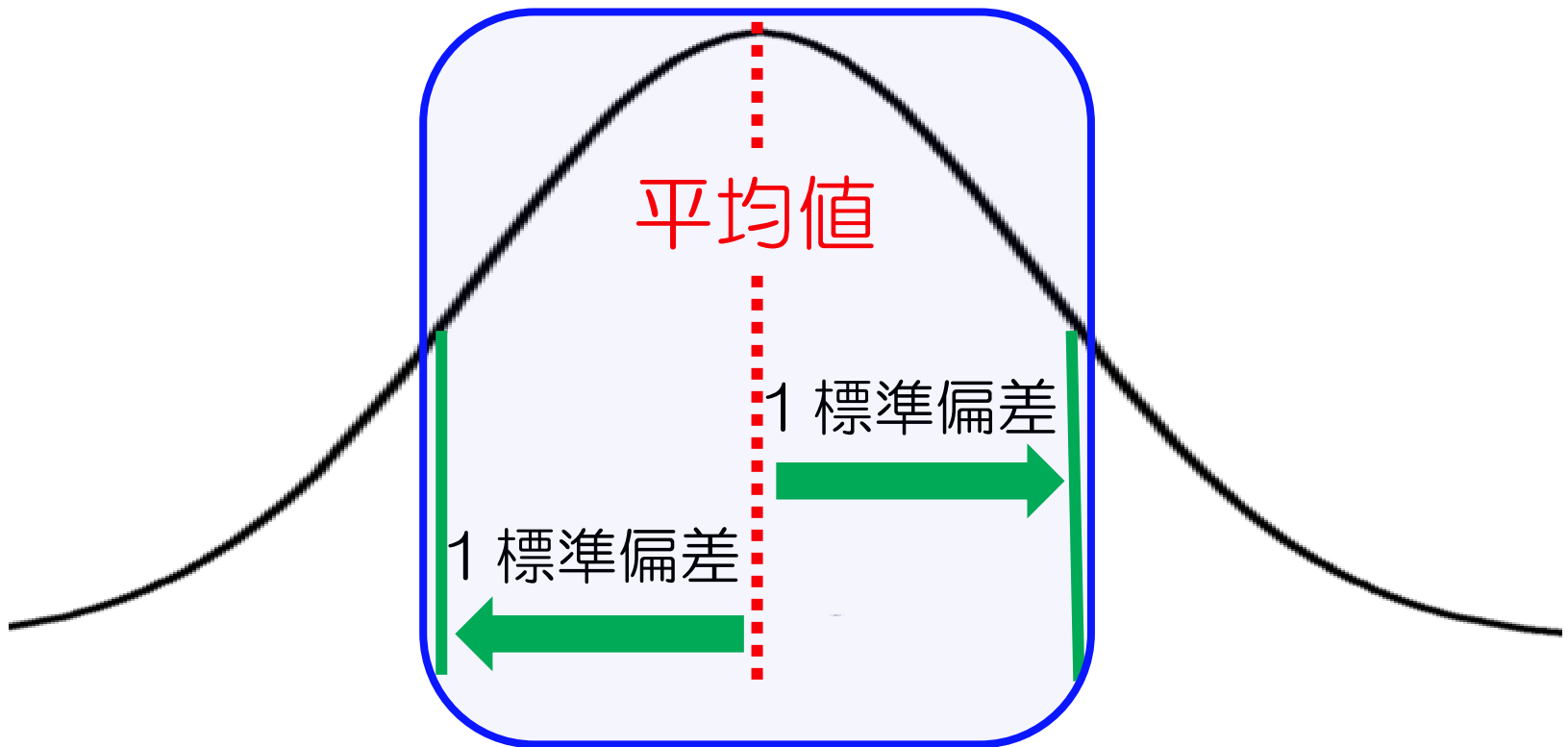
(ばらつき)



ここで大事なこと

平均値 $\pm$ 1 標準偏差の中に

全データの「68%」がある



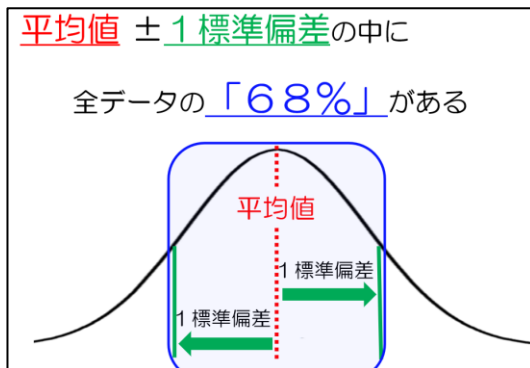
つまり

2	5	9	11	13
---	---	---	----	----



結論！

平均値 8点 ± 4点 は普通



(普通：全データの68%内にある)

平均値「8」

2	5	9	11	13
---	---	---	----	----



平均より下だから
成績は悪い？

平均より上だから
成績はいい？

この仮説は間違い！