

ST・SS 未来創造 I

「データサイエンス」統計の基本



A教室

白色光でテスト



B教室

昼色光でテスト



結果

B組がA組よりも平均点5点良かった

⇒ 「昼光色がテストに良い影響を与える」
と言えるか？



問い

「昼色光がテストに良い影響を与えた」と言えないとするならば、
どんな理由が考えられるか？

できるだけ多くの理由を考えて
みよう？

理由 1

もともとB組の方に成績の良い人が集まっていたのではないか？

⇒ 「無作為抽出法」で分ける
単純抽出法(「くじ」や「乱数表」を使う)
層別抽出法(男女など各層から抽出)



理由 2

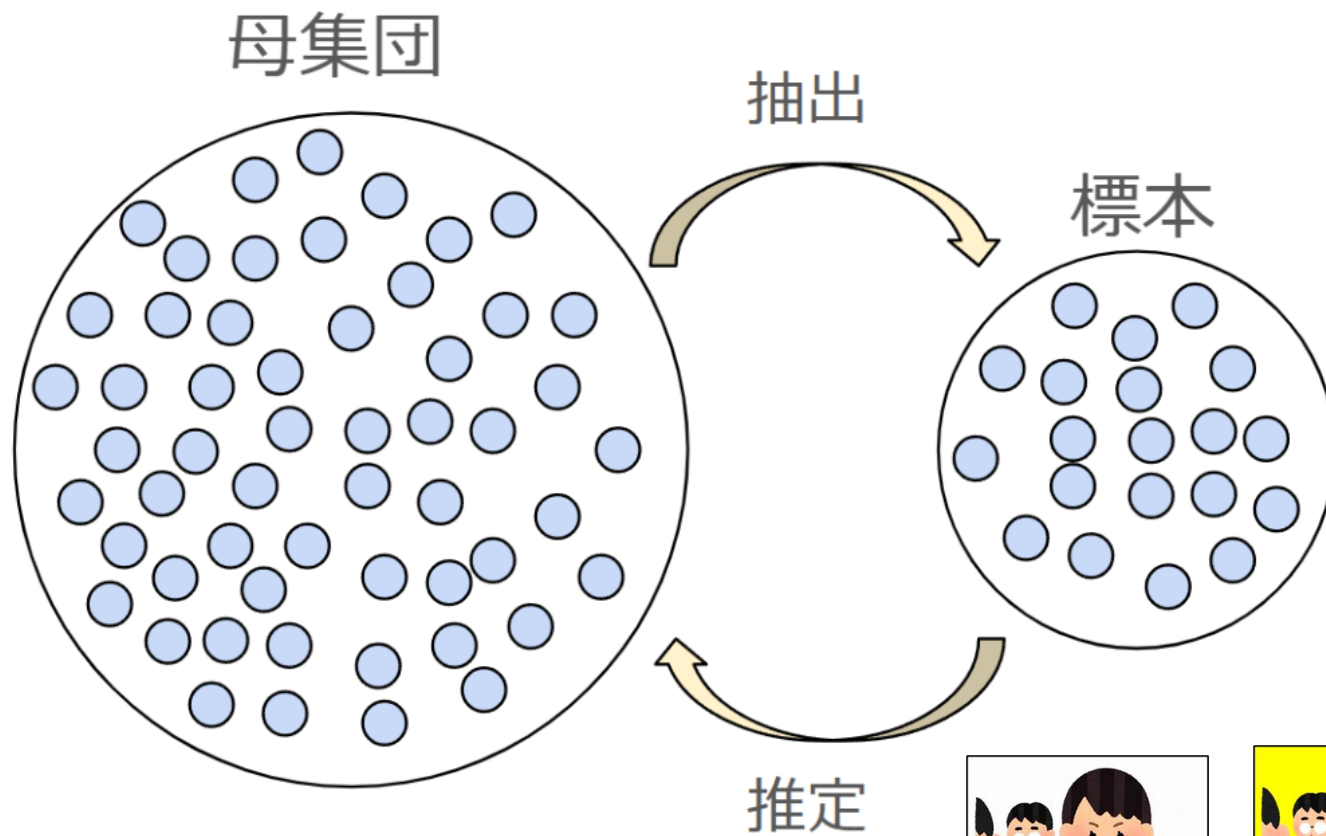
平均の 5 点差は、「照明」の影響ではなく、
他の影響が関係したのではないか？

例：「教室の室温の差」「騒音の差」「B組でたまたま同じ
ような問題をやっていた」・・・など

⇒ 「照明」以外の条件はすべて同じにするように
配慮する。

理由 3

平均の5点差は、全体から抽出されたA,B教室の生徒を対象にして、得られたデータであって、すべての生徒について適応するものではないのではないか



全ての生徒

A,B教室の生徒

理由 4

平均の 5 点差は、たまたま偶然生じたもので
光の色の影響は関係ないのではないか

理由 4

5点差はたまたま偶然生じたもの
光の色の影響は関係ないのではないか

⇒ 「5点差」が偶然か、めったに起きないものか？

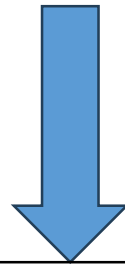
⇒ **有意差検定法** = その「差」が何か意味(理由や原因)がある「差」なのかどうかを判断する方法

⇒ 「両クラスに差がない」と仮定して(帰無仮説)
「5点差」がつく確率を調べる = P値

5 点差が生じる確率 = P値 (Probability確率)

⇒ 確率30% = P値0.3
確率10% = P値0.1

確率 5 % = P値0.05
確率 1 % = P値0.01



**偶然ではないと
判断する範囲**

コインを投げて

確率



表が3回続けて
出る

?



表が4回続けて
出る

?



表が5回続けて
出る

?

確率



表が3回続けて
出る

12.5%



表が4回続けて
出る

6.3%



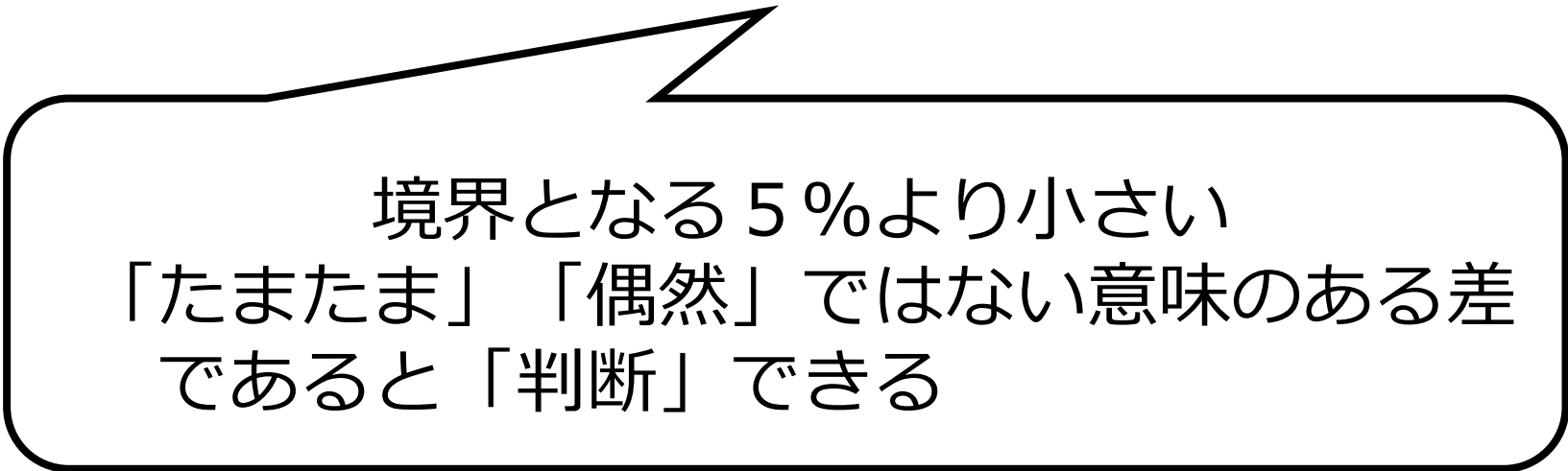
表が5回続けて
出る

3.1%

B組(昼色光)の平均が 5 点良くなる確率を調べる

⇒ P値は0.015であった。(計算の仕方は後述)

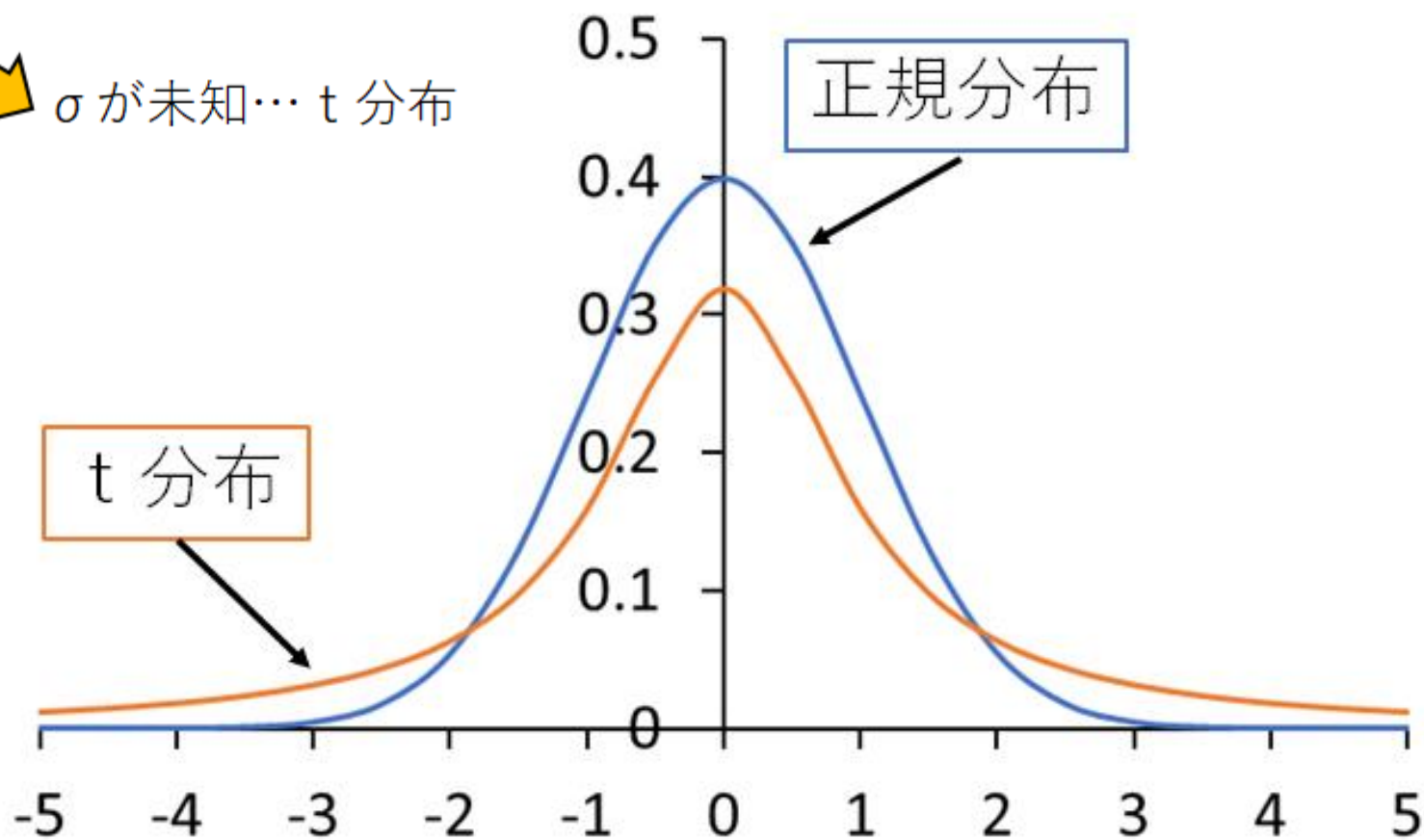
⇒ 平均が5点良くなる確率は**1.5%であった**



境界となる 5 %より小さい
「たまたま」「偶然」ではない意味のある差
であると「判断」できる

σ が既知...正規分布

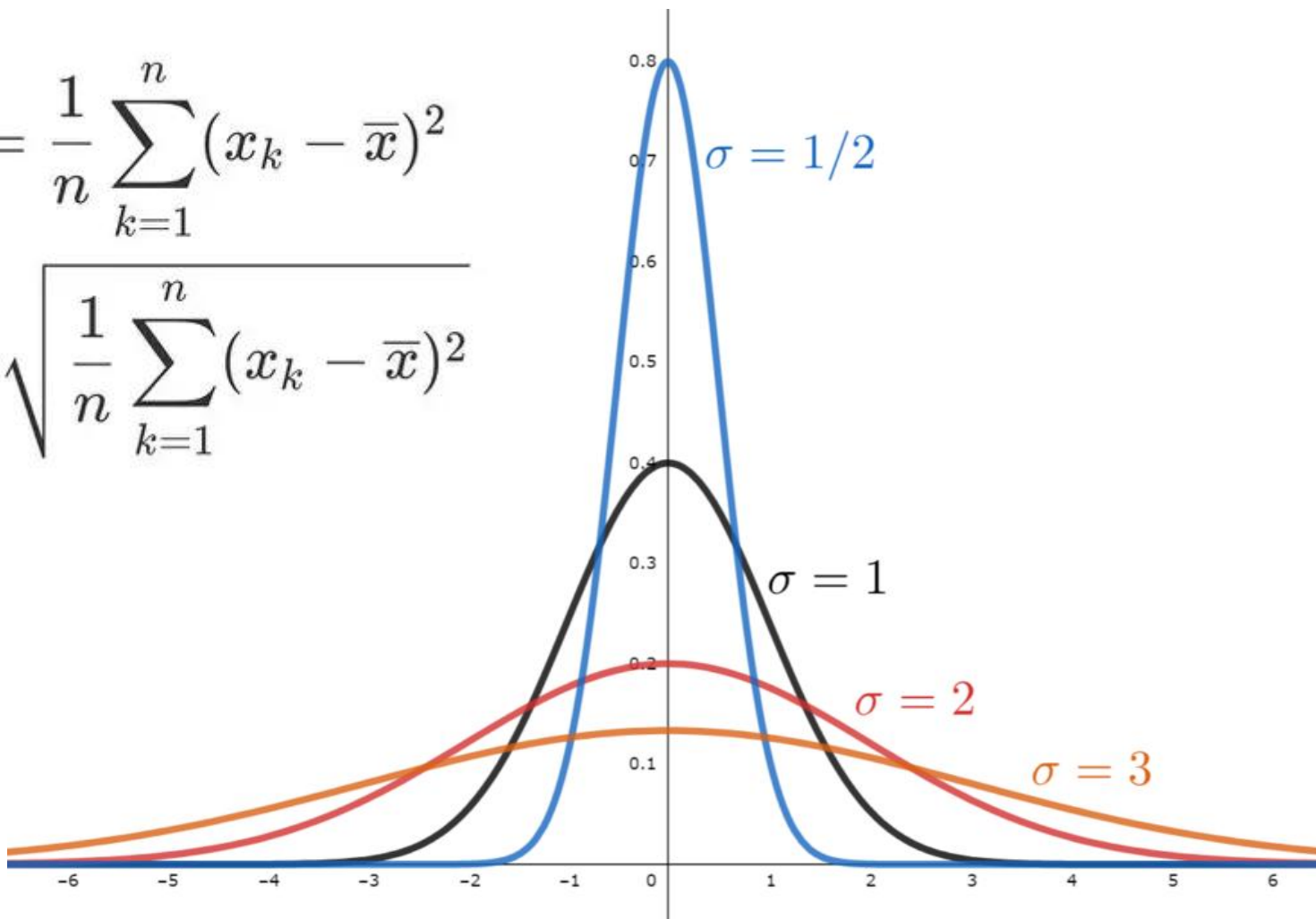
σ が未知... t 分布



σ = 散らばり度

$$\sigma^2 = \frac{1}{n} \sum_{k=1}^n (x_k - \bar{x})^2$$

$$\sigma = \sqrt{\frac{1}{n} \sum_{k=1}^n (x_k - \bar{x})^2}$$



「正規分布」 身長、テストの点数、製品の重量など



σ (標準偏差) = 1 とする

