

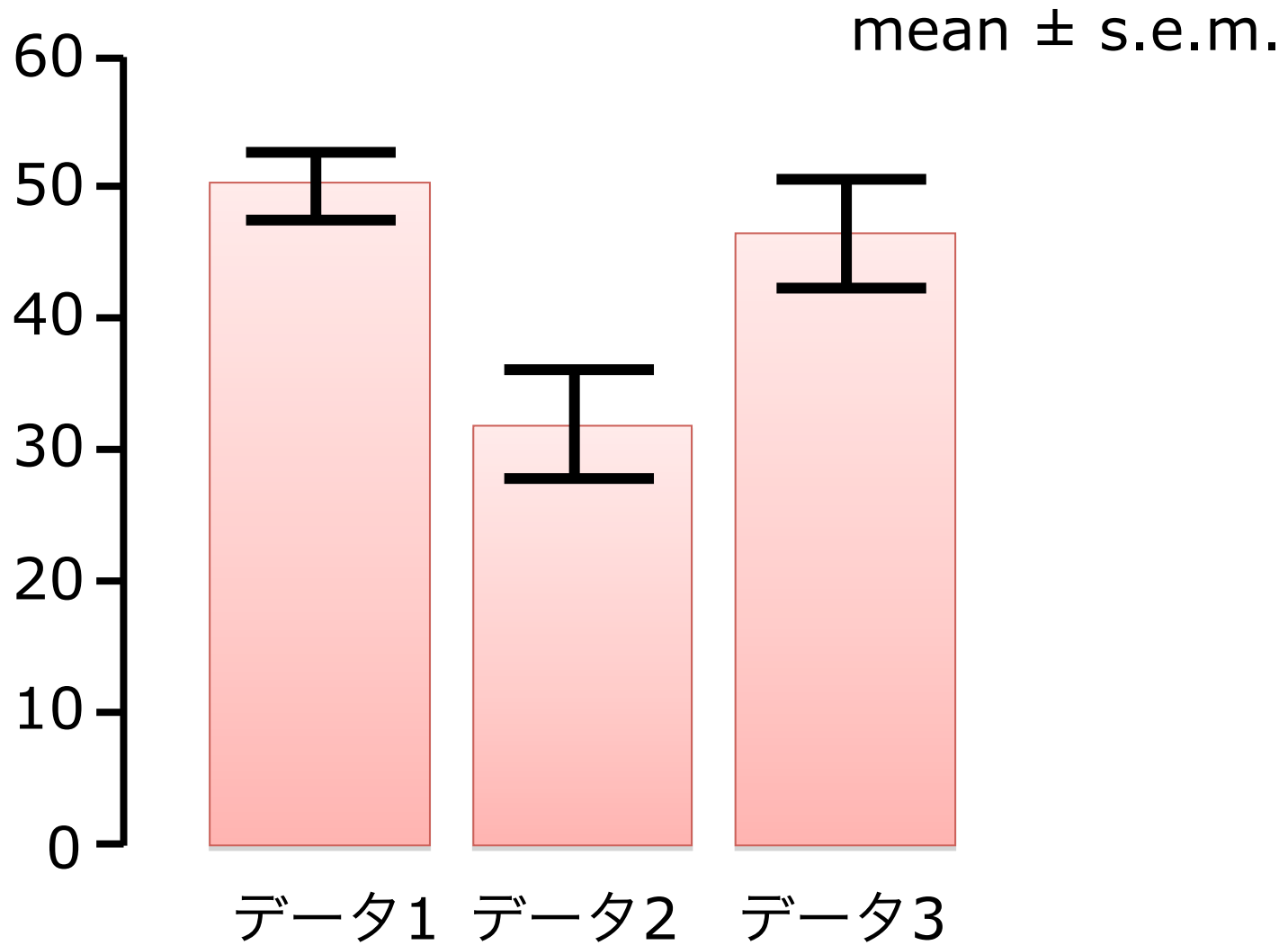
# 生体情報論演習

## - 統計法の実践 第5回 -

2011. 7. 8.

京都大学 情報学研究科 杉山磨人

# 先週の解答(棒グラフ)



# 2群の検定 (1/2)

- データ1 と データ2 の平均値を比較する
- 対応のあるt検定 (両側) を行うと,  
 $P = 1.22 \times 10^{-5} < 0.01$
- 結論: データ1 と データ2 の平均値には差がある
- 論文ではよく,  
「データ1 と データ2 の平均値には有意  
(significant) な差がある ( $P = 1.22 \times 10^{-5} < 0.01$ )」  
と書く

## 2群の検定 (2/2)

- データ1 とデータ3 の平均値を比較する
- 対応のある $t$ 検定(両側)を行うと,  
 $P = 0.48 > 0.05$
- 結論: データ1 と データ3 の平均値には差がない
- 注意: 「差がない」と言っははいけない, とする立場もある→「差があるとは言えない」
  - 実際には, この区別はあいまいで, あまり気にする必要はない

# 2群に対応があるときとないとき

- 新しいマウス用のえさを開発した. このえさを食べることで, マウスがやせることを示す
  - 生まれてから古いえさを食べ続けているマウス群と, 新しいえさを食べ続けているマウス群の体重を比較する
    - 2つのデータは独立している(対応がない)
  - あるマウス群にまず1ヶ月間古いえさを食べさせて体重を測定し, その後新しいえさを1ヶ月間食べさせて体重を測定する
    - 2つのデータは対応している(独立でない)

# データの例

独立なデータ

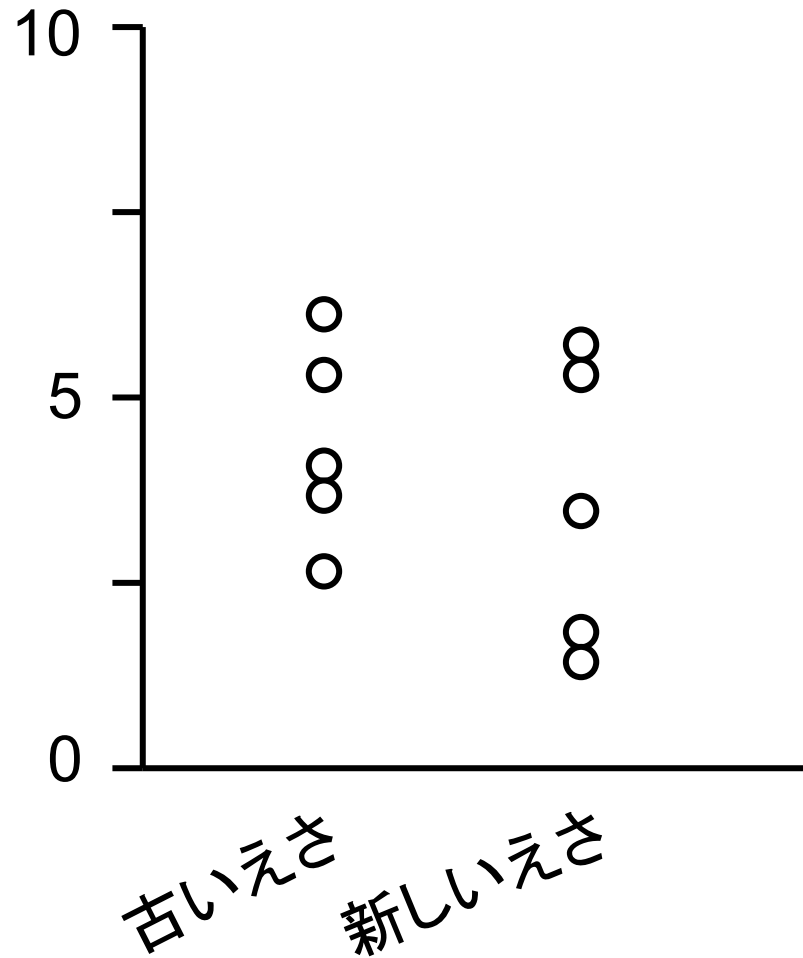
古いえさ	新しいえさ
5.3	3.5
4.2	5.7
4.4	2.3
2.8	2.1
6.3	5.3

対応のあるデータ

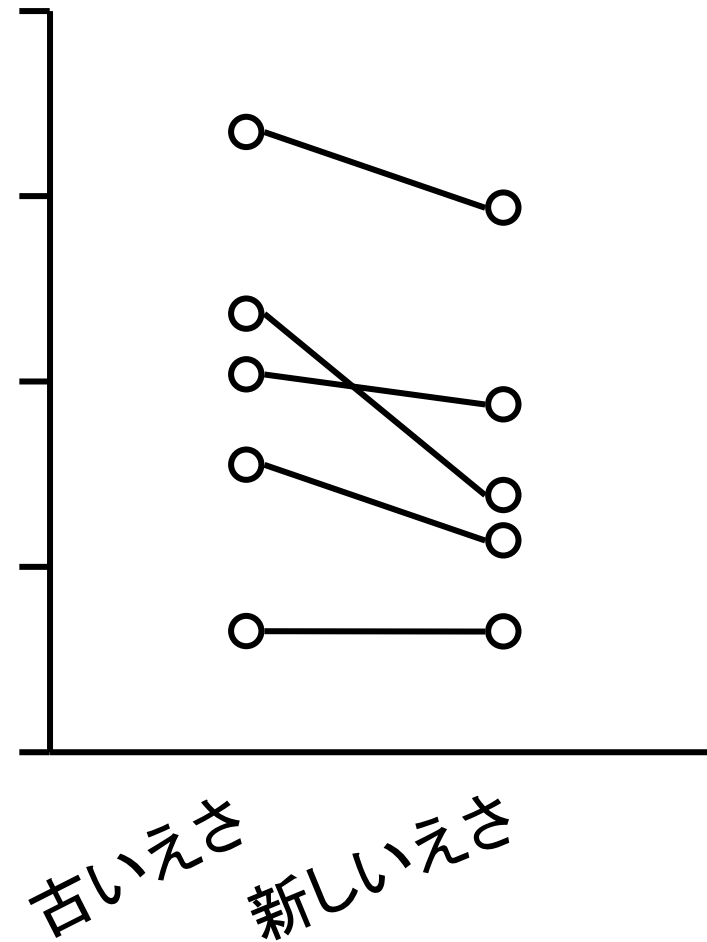
古いえさ	新しいえさ
5.1	4.8
5.5	3.3
8.6	7.4
2.1	2.4
4.2	3.1

# グラフ(散布図)の例

独立なデータ



対応のあるデータ



# どの $t$ 検定を使うのか

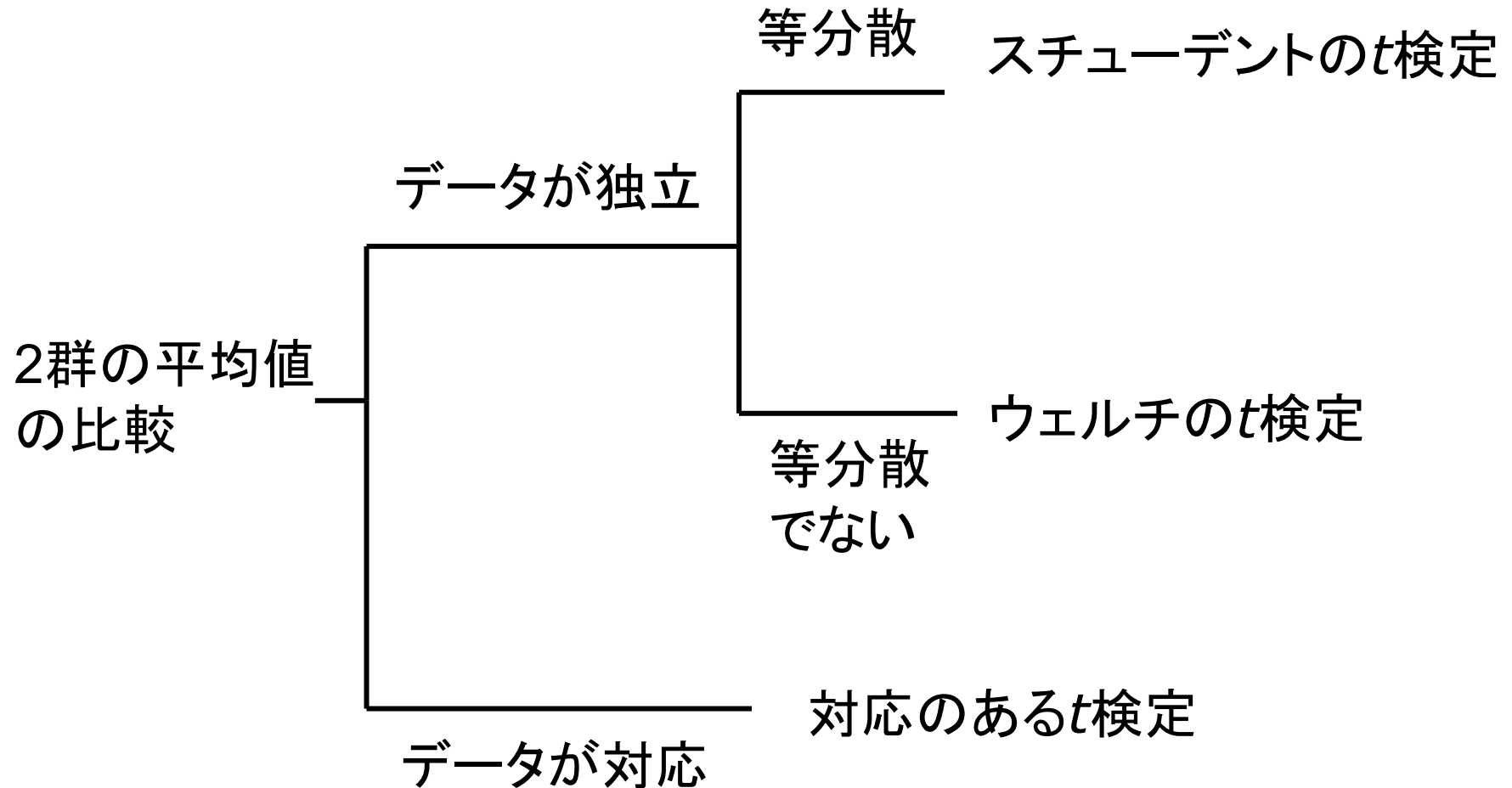
- 独立した2群の平均値を比較する  
→ 独立な $t$ 検定 (いわゆる $t$ 検定) を使う
- 対応のある2群の平均値を比較する  
(先週やった)  
→ 対応のある $t$ 検定 を使う



# 独立な $t$ 検定にも2種類ある

- 2つの群の分散が等しいのかどうか
- 等分散を仮定する $t$ 検定
  - スチューデントの $t$ 検定
- 等分散を仮定しない $t$ 検定
  - ウェルチの $t$ 検定
- 分散が等しいデータはほとんどない
  - ウェルチの $t$ 検定が使いやすい
- ただし, どちらを使ってもほぼ結論は同じ

# t 検定のまとめ



# 両側検定と片側検定

- どの検定を使うときにも、両側検定と片側検定の2種類がある
- とにかく「差がある」ことを示したいときには両側検定を用いる(先週はこっち)
- 「大きくなる」や「少なくなる」など、どちらか一方のみを示したいときには片側検定を用いる
- 生命科学では、片側検定を使うべき状況が多いが、間違っって両側検定を使っている論文がたくさんある

# 課題

- 実験1
  - 新しい遺伝子*NGene*を発見した. この遺伝子を破壊したゼブラフィッシュの体長が短くなるかどうかを調べる.
  - 破壊したゼブラフィッシュ20匹と, 破壊しなかったゼブラフィッシュ20匹の体長を比較する.
  - 遺伝子*NGene*を破壊すると, 体長は短くなるのか?
- 実験2
  - 新しいえさをつくった. このえさを食べたゼブラフィッシュの体長が短くなるかどうかを調べる.
  - まずゼブラフィッシュ20匹の体長を測定する
  - その後, その20匹にえさを1ヶ月食べさせ, 再び体長を測定する
  - えさを食べると体長は短くなるのか?

# 課題

- 実験1, 2のエラーバー付き棒グラフをそれぞれ描く
- 実験1, 2をそれぞれ検定する
  - どの検定法を用いるのか書く(データが独立か対応してるかに注意)
  - $P$  値を書く(片側検定か両側検定かに注意)
  - 結論を書く
- 来週までに提出