

生体情報論演習

- 統計法の実践 第6回 -

2011. 7. 14.

京都大学 情報学研究科 杉山磨人

先週の課題の解答

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?

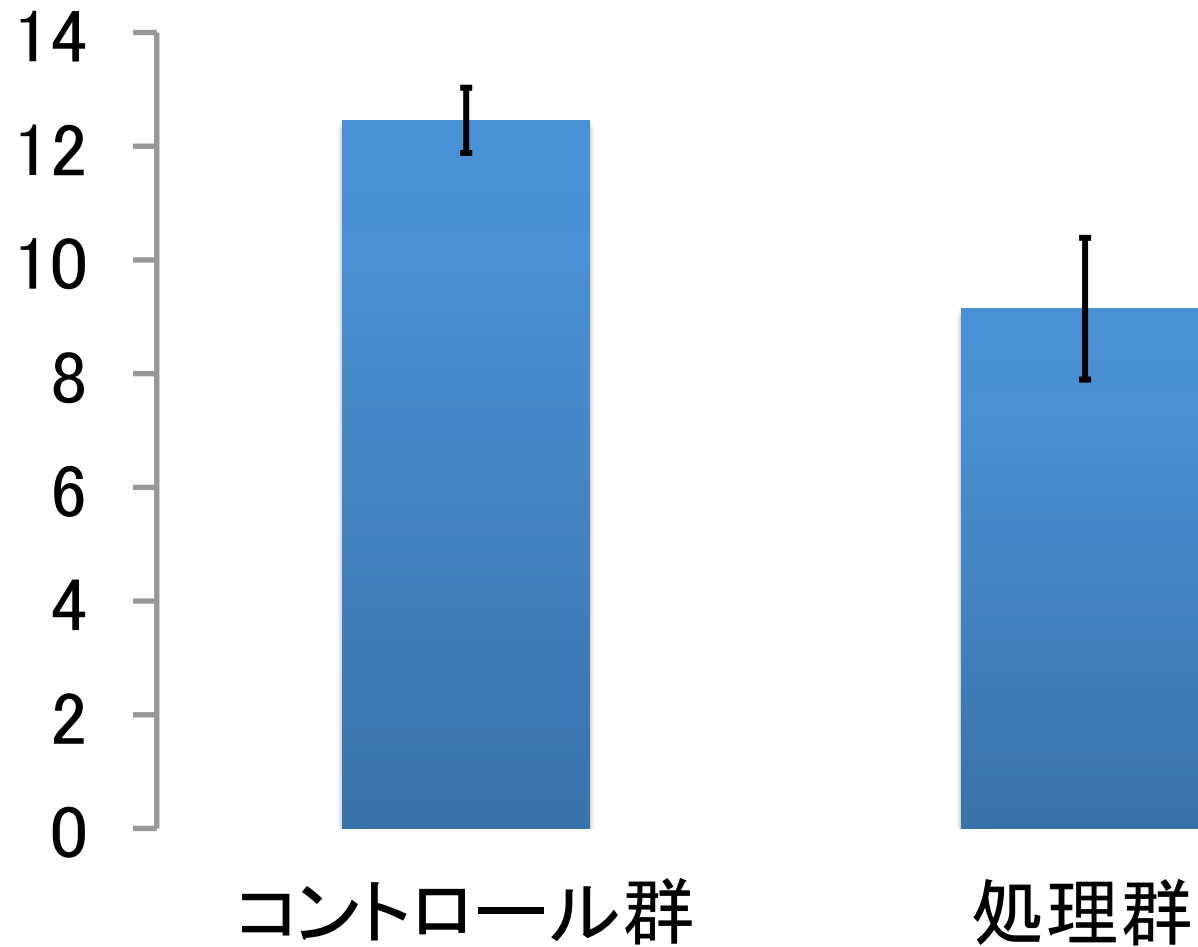
- 実験2

- 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のエラーバー付き棒グラフを描く
- 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

DATA1の棒グラフ



実験1

- 「ウェルチのt検定」を使う
 - データは独立している
 - 分散は同じではなさそう
- 片側検定
 - 体長が短くなるかどうかだけを見たい
 - 体長が長くなっても意味がない
- $P = 0.011463586 > 0.01$

実験2

- 「対応のあるt検定」を使う
 - データは対応している
- 片側検定
 - 体長が短くなるかどうかだけを見たい
 - 体長が長くなつては意味がない
- $P = 0.008843812 < 0.01$

実験1と2を比べると・・・

- データは同じなのに、1では有意な差がなく、2では有意な差がある（結果が正反対）
→2のほうがわずかな差を検出できた
- 一般に、対応のあるデータのほうが、有意な差を見つけやすい
- しかし、対応のある場合は「順序効果」を考慮しなければならない
 - コントロール群の実験と処理群の実験、どちらを先にするのか？
 - 結果が変わる場合があり、実験が難しい

結論(ガイドライン)

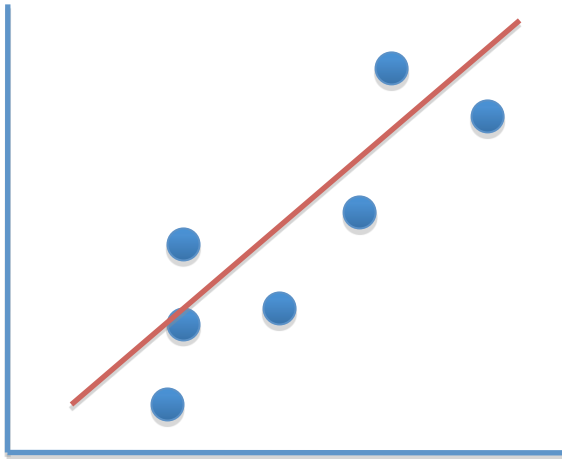
- データがたくさん(~10以上)とれるときには独立な実験にする
- データが少ないときには、対応のある実験にする
 - ただし、順番を変えて実験をおこなうなど「順序効果」の対策をしなければならない

回帰分析

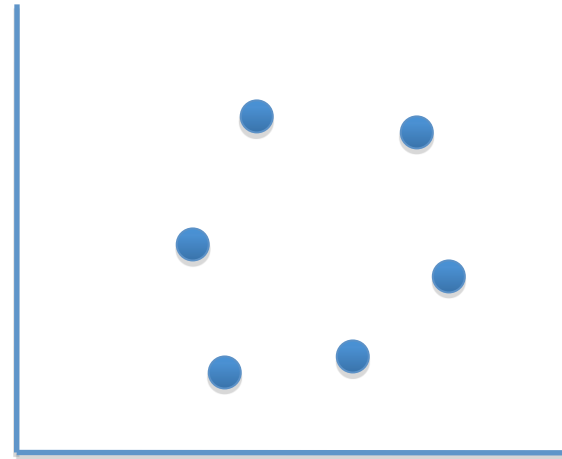
- 「差があるかどうか」ではなく、2つの変数間の「**相関**」を調べたい
 - 例1: 身長が高くなると、体重が増える？
 - 例2: 雨が降ると、作物が育つ？
 - 例3: 薬の量を増やすと、血圧が下がる
- **回帰分析**という方法を使う
 - xという変量とyという変量があったとき,
$$y = ax + b$$
という式のaとbを求める
 - 直線関係(比例関係)しかわからないことに注意

データの例

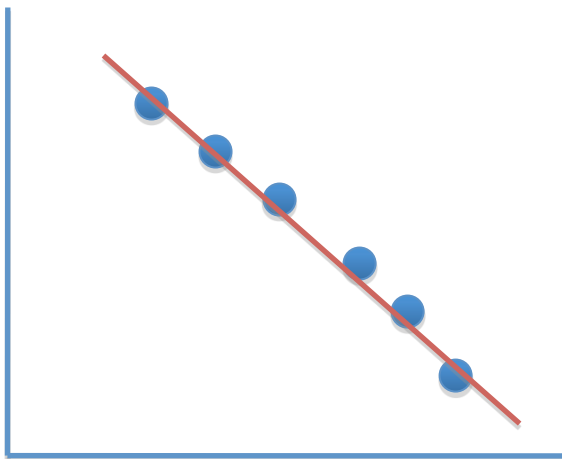
○



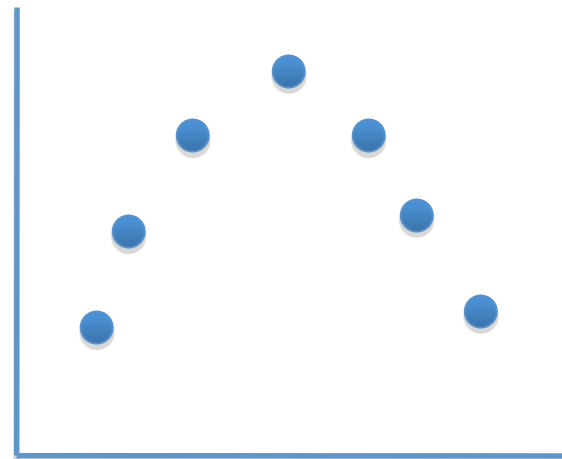
×



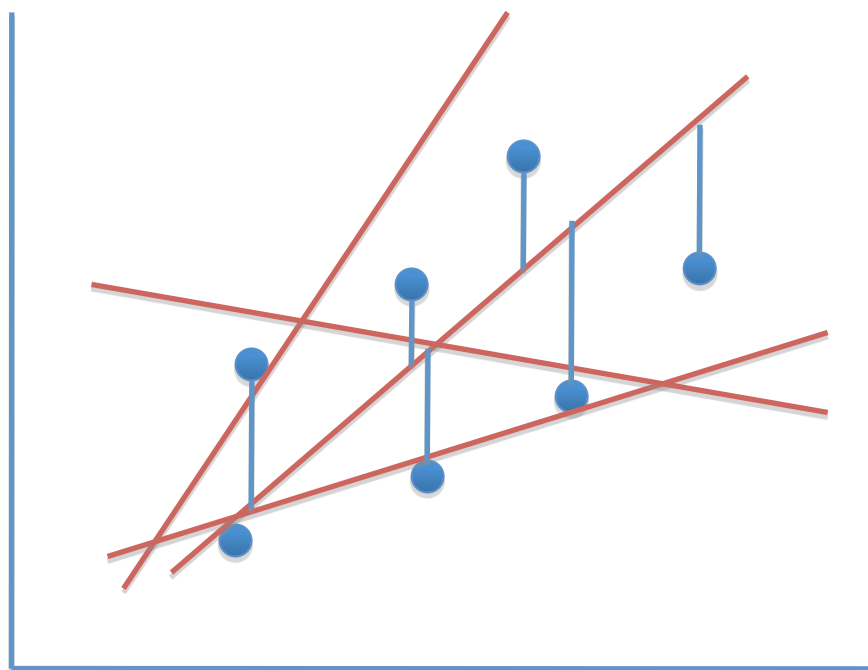
○



×



どんな式を求めるか



最小二乗誤差が
最小になる直線を
求める

課題

- データに対して回帰分析をやる
 - 式を求める
 - グラフも描く
- 来週までに提出