植物生態学特論 II 代講

全部で 2 回講義の 2

「検定」の使われかたを観察してみる

— 「検定」ってそんなにエラいのか?

http://hosho.ees.hokudai.ac.jp/~kubo/stat/2005/

講釈: 久保拓弥 kubo@ees.hokudai.ac.jp

この 2 回だけの統計学授業でやること

- ○ 自然科学の データ解析 に統計学は必要不可欠
- ○ しかし大半のユーザーは 何もわからん 状態で使ってる
- ○ この授業の目的はその「わからん度」を少しだけ下げる こと

- **第 1 回**: 2005.07.04 (月)
  生態学研究で得られたデータを解析するための統計モデリング
  — 理解できる統計学めざして
- **第 2 回**: 2005.07.06 (水)
  「検定」の使われかたを観察してみる
  — 「検定」ってそんなにエラいのか?

個別的なワザより全体に共通する考えかたを— ただし内容は偏ってるよ

今日のハナシ: 検定とは何か? エラいのか?

- 論文とかのデータ解析で「検定」がよく出る
- あれは何をやってるのか? 「ゆーい差」って何か嬉しいのか?

(目次)

1. 検定って何なの

「肥料は花数ふやす」という架空実験の例で考えてみる

2. 検定はそんなにエラいのか ?!

そのあたり，検討してみましょう

3. 検定つかうときの注意点

こまごまとしたつけたし

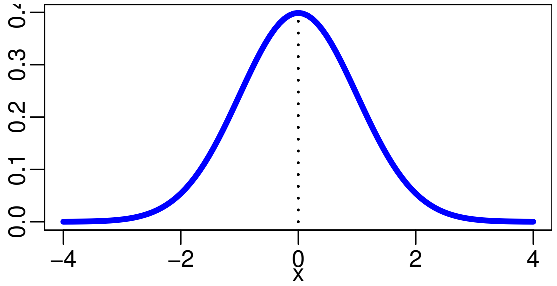
「肥料は花数を増やすのか」問題で考える

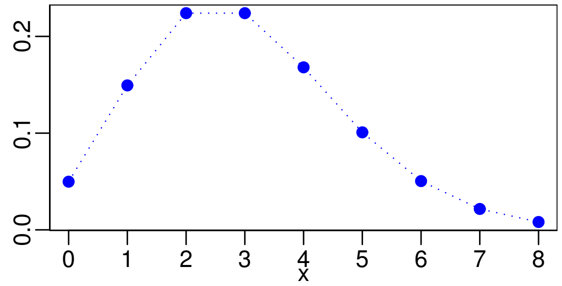
1. 検定って何なの

万能なる Parametric Bootstrap 法!

(前回の復習) 乱数 (標本) と推定の関係

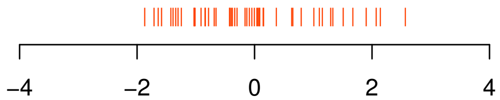
- モデル
- 確率分布
- 母集団

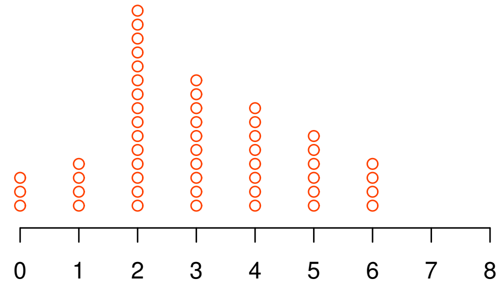




サンプリング ⇓ ⇑ (パラメーター) 推定

- データ
- 乱数
- 標本集団



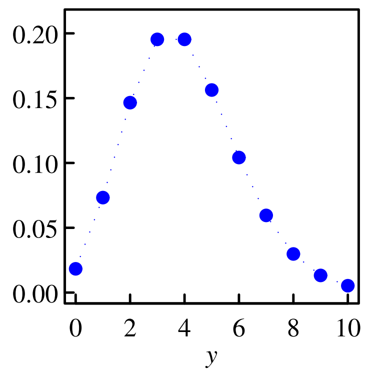


- 自然科学者は何か ばらつきのある自然現象をみたときにそれが確率論的モデルによって生成された，と仮定する → モデルによる単純化
- このばらつきのあるデータから確率論的モデルのカタチを特定してやることがパラメーター推定である → モデル選択や検定につながる

(前回の復習) じゃあ推定ってのは何なの?

ポアソン分布の推定

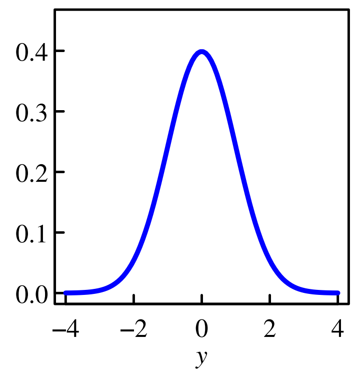
5 4 3 2 4 2 4 1 7 1 ⟶



正規分布の推定

```
 1.4851004 -0.9912880 -0.1092131
-2.1752314 -0.3779424  1.1360432
 1.2493592 -1.2405408 -0.4425550
```

⟶

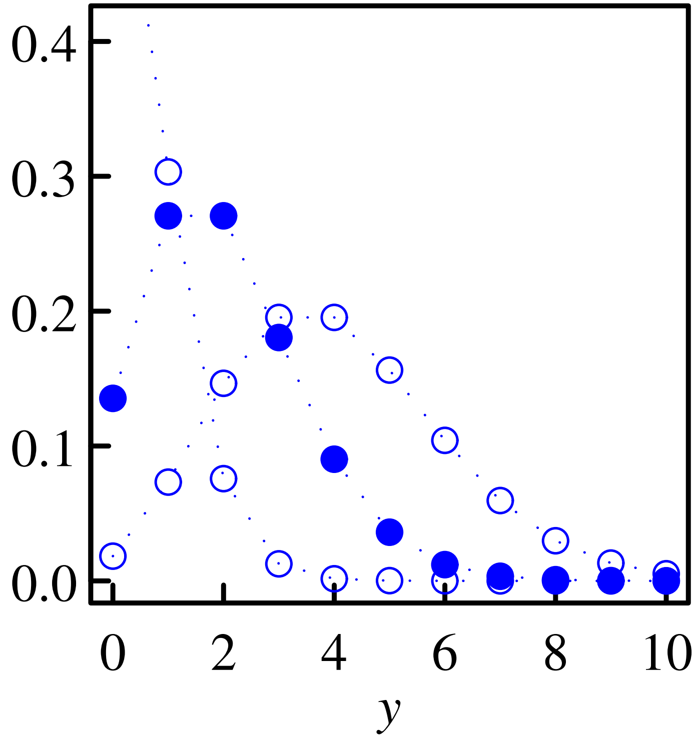


乱数とみなされる標本集団
⟶ 母集団すなわち**確率分布**を決め
そのパラメーターを決めてやる技法

ポアソン分布 (Poisson distribution)

- 離散分布 $y_i \in \{0, 1, 2, \cdots, \infty\}$

R の関数: `dpois(`y, λ`)`



- 確率密度関数 (paramter: λ)

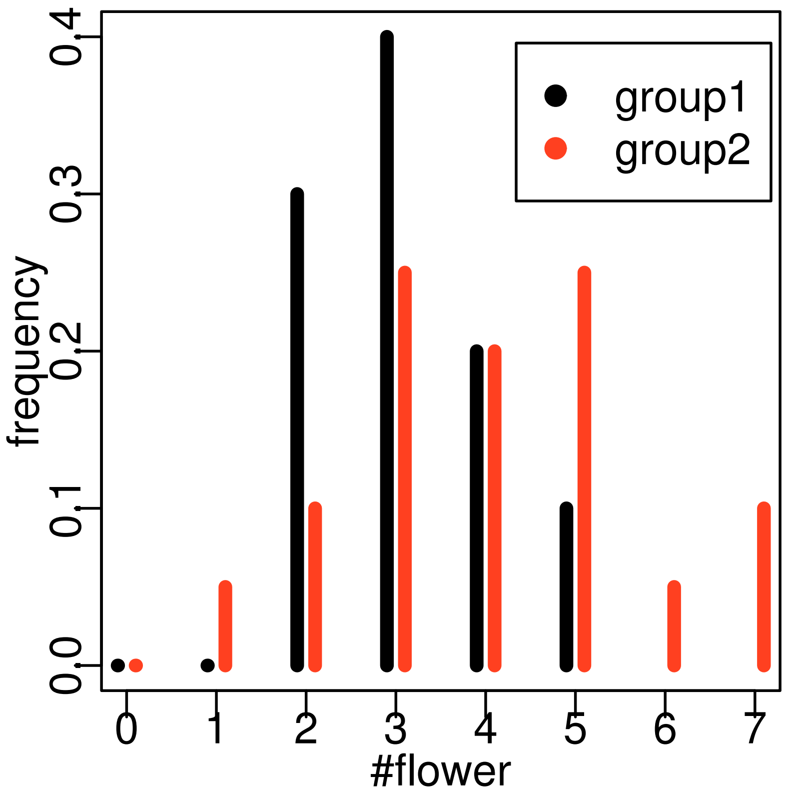
$$\frac{\lambda^y \exp(-\lambda)}{y!}$$

- 期待値 λ，分散 λ

- 使いどころ: 「一定時間にかかってくる電話の回数」……上限を設定できないカウントデータ
  - 産卵数・種子数

- 個数のデータが得られたら，まずは「ポアソン分布で説明できないか?」と考えてみる

さて今日の架空実験: 肥料処理と花の個数



[実験]

- 集団 1 — 10 個体
  - 肥料なし
- 集団 2 — 20 個体 (!)
  - 肥料あり

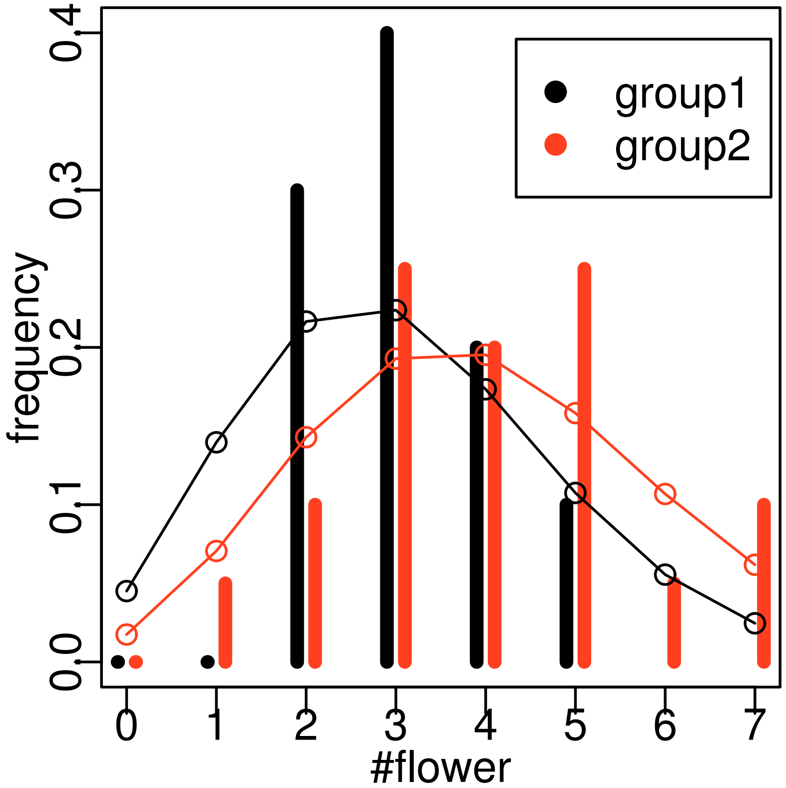
[調べたいこと]

- 肥料やったら花数は「増える」のか?

[実験結果] 花数 (上の図も参照)

集団 1 (平均 3.10) 5 4 3 2 2 3 3 3 2 4
集団 2 (平均 4.05) 2 4 4 3 2 6 3 5 5 3 4 5 3 3 7 1 7 5 4 5

(またしても) 「神」の立場なら知っていること



[実験]

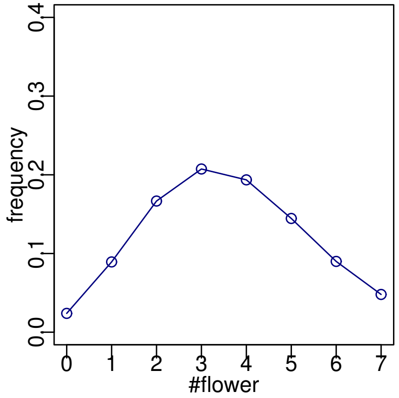
- 集団 1 — 10 個体
  - 肥料なし
  - 平均花数 3 のポアソン乱数
- 集団 2 — 20 個体 (!)
  - 肥料あり
  - 平均花数 4 のポアソン乱数

- 「ホントは」肥料をやると花数が増える
- 観測者たる人間の立場でそれを知ることができるか?

検討の出発点: 現象を説明できそうなモデルは?

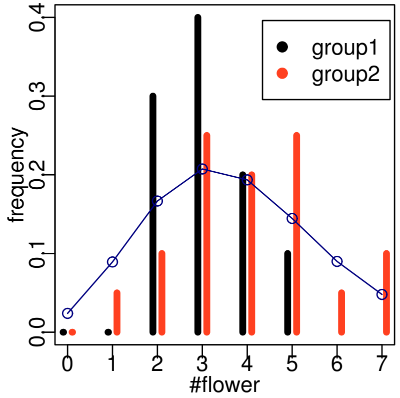
1. 「肥料は花数を増やさない」モデル (帰無仮説) — 単純モデル

モデル (確率分布)

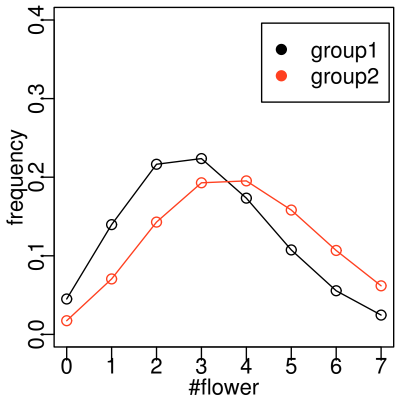


→

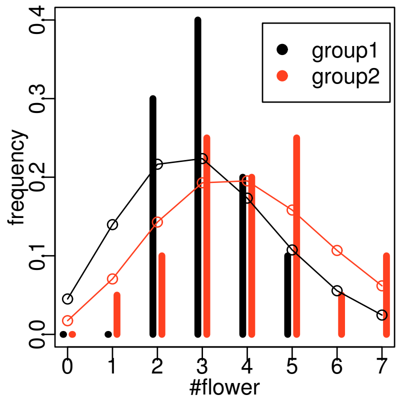
乱数 (標本)



2. 「肥料は花数を増やす」モデル (対立仮説) — 複雑モデル

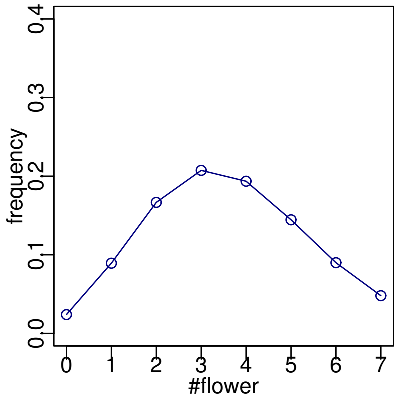


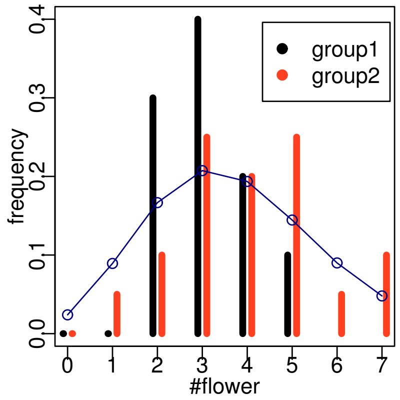
→



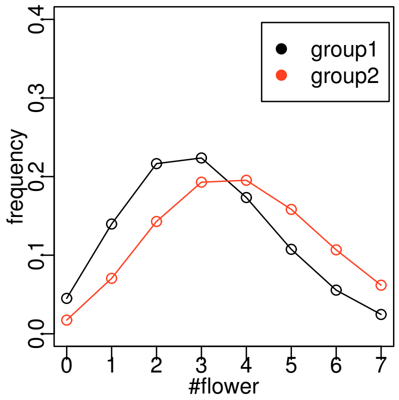
説明できそうなモデルたちの正体は ……

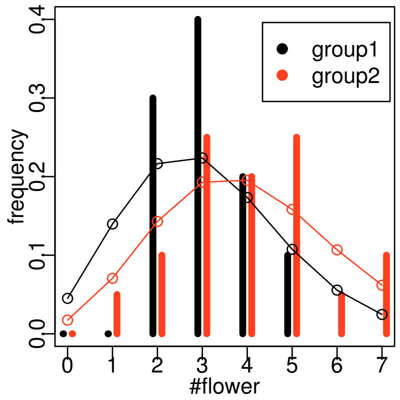
1. **帰無仮説**: 平均 3.73 のポアソン分布から得られた $\{10, 20\}$ 個の標本





2. **対立仮説**: 平均 $\{3.10, 4.05\}$ のポアソン分布から得られた $\{10, 20\}$ 個の標本



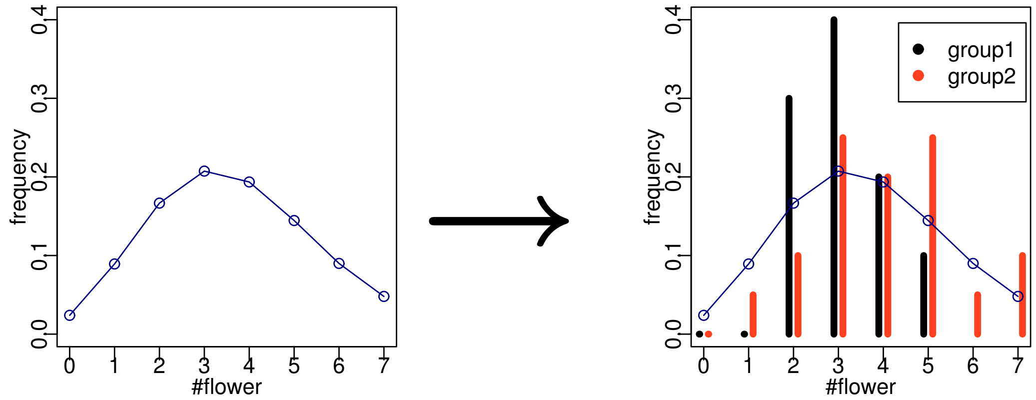


「肥料で花数増加」という予断がある（のでこういう実験やった)

(統計学的)「検定」の不思議さ: 帰無仮説だけを?

- 検定では 帰無仮説だけを重視する
- しかも 帰無仮説がまちがってる (現象を説明できん)
  「可能性」だけを調べる (!)

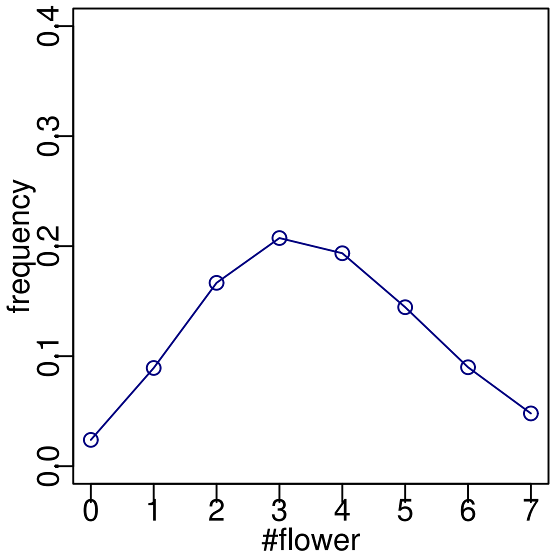
1. 「肥料は花数を増やさない」モデル (帰無仮説)



2. 「肥料は花数を増やす」モデル (対立仮説) → (そんなの知らねーよ)

- その「したごころ」は ?!
— 帰無仮説が「まちがってる」なら 対立仮説が正しかろう

帰無仮説で観測された現象は説明できるのか?



⟶

[得られた観測値 (花数)]

集団 1 (**平均** 3.10; **個体数** 10)

花数: 5 4 3 2 2 3 3 3 2 4

集団 2 (**平均** 4.05; **個体数** 20)

花数: 2 4 4 3 2 6 3 5 5 3

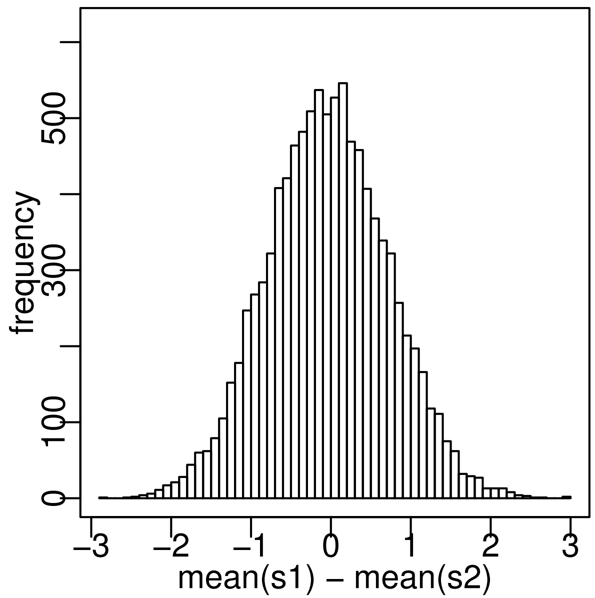
4 5 3 3 7 1 7 5 4 5

ここで集団 1 と 2 の「平均値の差」に注目する

- **観測された「差」**: 3.10 - 4.05 = -0.95 (**検定統計量**)
- **帰無仮説から得られる 10 & 20 個体集団で「差は -0.95」は よくあることなのか? めったにないことなのか?**

「帰無仮説な植物」で実験を 10000 回やってみる

この原始的な解決方法が Parametric Bootstrap 法



- 平均 3.73 のポアソン分布から乱数を
  - ニセ集団 1 のために 10 個
  - ニセ集団 2 のために 20 個

  発生させる
- 「差」 = (ニセ集団 1 の平均) - (ニセ集団 2 の平均)
- これを 10000 回くりかえす (左の図)

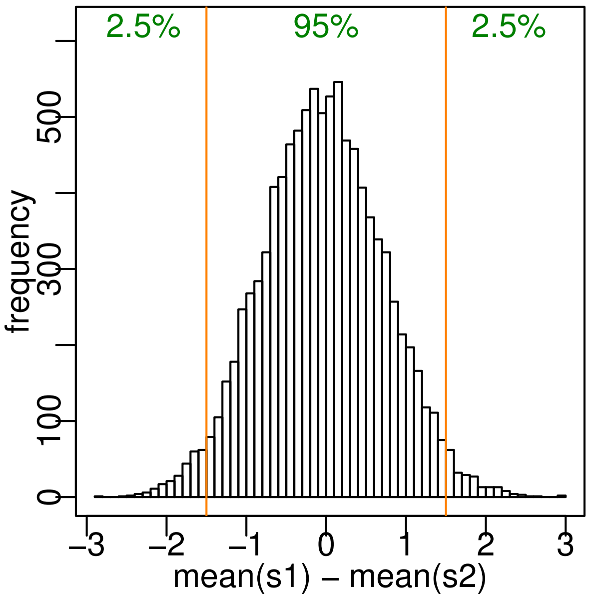
[Parametric Bootstrap 法の特徴]

- R まかせ，ややこしい数式は不要
- どんな確率分布にも使える (「このデータはどういう確率分布?」をよく見る)

「花数は肥料に無関係」で説明が苦しい領域

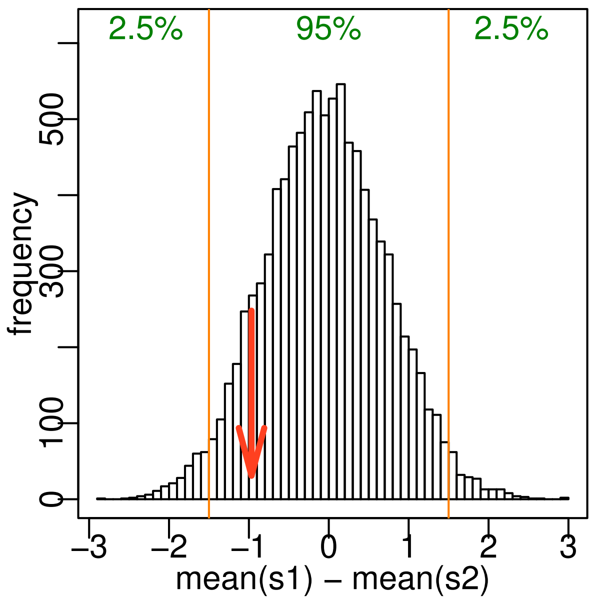
「帰無仮説が正しい」と仮定して得られた「差」の分布

めったにでない (5% 未満) 「差」ってのはどれぐらいか?



- めったにおこらない確率を危険率という (ここでは 5% — 10000 回のうち 500 回の部分)
- 「危険」な「差」からなる部分を棄却域という
- 左の図の左右 2.5%が棄却域
  「めったにない差」(帰無仮説が正しいなら)
- 危険率は 5% でなくても構わない (てきとー)
- とゆーか，5% とかに意味ナシ (自然科学者としては)

検定の結論: 「肥料やれば花数増加」とは言えん



[得られた観測値 (花数)]

観測された「差」: 3.10 - 4.05 = -0.95

集団 1 (**平均** 3.10; **個体数** 10)

花数: 5 4 3 2 2 3 3 3 2 4

集団 2 (**平均** 4.05; **個体数** 20)

花数: 2 4 4 3 2 6 3 5 5 3 4 5 3 3 7 1 7 5 4 5

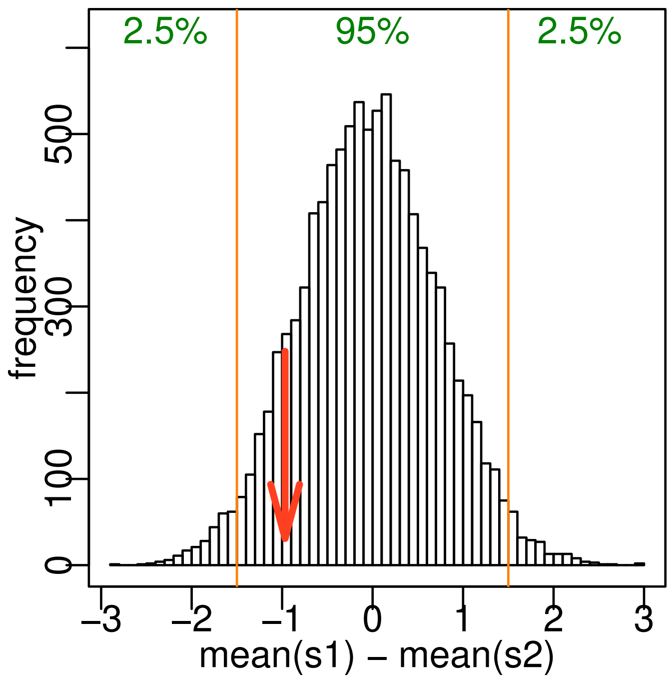
- 「差は -0.95」は「めったにないこと」ではない
- したがって 帰無仮説「肥料やっても花数は変化しない」を棄却できない → オモテがでる確率に統計学的な有意差はなかった
- (もし実験結果が差 -1.5 とかだったら) 帰無仮説は危険率 $< 5\%$ で棄却できる → オモテがでる確率に統計学的な有意差があった

2. 検定はそんなにエライのか

そのあたり，検討してみましょう

検定についてまわる二つの過誤

↓帰無仮説が	「めったにない差」(帰無仮説棄却)	「よくある差」(棄却できない)
正しいとき	第一種の過誤	OK
正しくないとき	OK	第二種の過誤



- **第一種の過誤** (やらかす確率 5%)
  - 真実: 肥料に花数増加の効果ナシ
  - 検定結果: 「肥料で花数が増えます」
- **第二種の過誤** (「検出力」の不足)
  - 真実: 肥料で花数が増加する
  - 検定結果: 「肥料は役たたずだ!」

検定の良いとこ悪いとこ

↓帰無仮説が	「めったにない差」(帰無仮説棄却)	「よくある差」(棄却できない)
正しいとき	第一種の過誤	OK
正しくないとき	OK	第二種の過誤

- 良い点 — 第一種の過誤をきちんと評価する
  - 帰無仮説を棄却するときに危険率を計算している
- 悪い点 — 第二種の過誤については直接的には検討しない

(この例題での) 第二種の過誤: 「肥料で花数が増えてるのに『肥料は役立たず』と結論してしまう」
これが検討されてない, ってどういうこと?

検定の非対称性

(前のペイジからのつづき)

- 悪い点 — 第二種の過誤については直接的には検討しない
  - 間接的には考慮する場合がある
  - 直接的には **検出力**を評価しなければならない

    → 実験計画法 (標本数を決める技法)

「帰無仮説だけ重視」という検定のポリシーによって

– **帰無仮説**が棄却できる場合

→ 危険率こみで「対立仮説」を採択できる

– **帰無仮説**が棄却できない場合 → **何も言えない**

なぜこんな「非対称性」あっても OK になってるのか?

なぜ「非対称性あってもいーや」と放置?

問: なぜ第一種の過誤 (帰無仮説の誤棄却) の危険性だけを計算?

答: 「肥料やったんだから花数が増えてあたりまえ」
という予断があるから
— 帰無仮説さえつぶせばいい (おかげで計算が簡単になる)

つまり計算が簡単 (ハナシが単純) になるからという理由で
「帰無仮説 (の棄却) だけ重視」ポリシーをとっている

(その前提?)

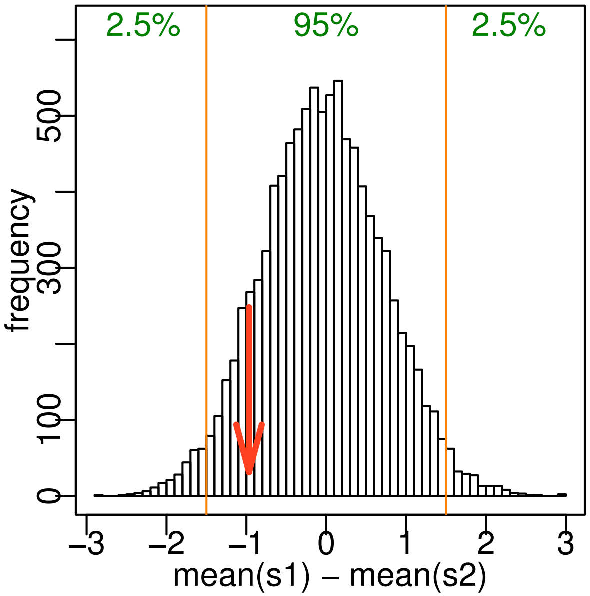
- 帰無仮説「処理と無処理の間に差がない」
  — 「処理をしたんだから無処理と違っていて当たりまえだ」
- 帰無仮説を棄却できなかったら実験は失敗なんだから,
  「何も言えない」でもさしつかえない (なげやり)

検定の利点・欠点を理解したうえで……

3. 検定つかうときの注意点

ところでホントに「検定」必要ですか?

検定の p 値って何ですか?



観測された「差」: **-0.95**
だとすると
この「差」がどれぐらい「めったにないこと」なのかを確率であらわしたもの
「差」が -0.95 なら $p = 0.2324$
(左の図のもとデータから計算できる)

……つまり

- 観測された「差」が「帰無仮説ただしい」わーるどでどれぐらいまれなことか，を確率で表現している
- つまり「捨てたい」帰無仮説 (だめ仮説) が「どれぐらいだめか」という数量

検定を使ったときに気をつけること

- 検定の説明について
  — どういう方法で検定をしたのかきちんとしめす
  - どういうモデルを考えているか，検定統計量は何か
  - 標本数 (あるいはこれをどうやって決めたか)
- 検定において標本数の大小はたいへん重要!!
- さらに……
  - p 値の大小は必ずしも「差の大きさ」をあらわすものではない
  - 「統計学的有意差」は必ずしも「生物学的有意差」ではない

このあたり理解してないとわけわからん
「ゆーい差」決戦主義者になってしまう

この本に「記録」されてる恐ろしい会話

生物学を学ぶ人のための統計のはなし
— きみにも出せる有意差 —

- 著者: 粕谷英一 (九州大・理・生物)
- 出版: 文一総合出版
- サイズ: A5 判 / 199 ペイジ
- ISBN: 4-8299-2123-4
- 発行年月: 1998.3

「この【ピンク本】を読まずにすますことはできない」
(三中信宏・書評 on “BK1” http://www.bk1.co.jp/)

なんてリアリティのある……!!

粕谷 “ぴんく本” 「統計のはなし」第一章「検定って，何?」より

浦井 飯山さん，村田が実験がいちおう終ったんで結果を検定したいそうなんですが，有意差が出る検定っていうのがあったと思うんですが，教えてください．

飯山 有意差の出る検定? — (中略) — いつでも有意差のでる検定なんてないよ．

浦井 村田，飯山さんは知らないらしい．誰か他の人に聞かなきゃだめみたいだ．

浦井君の「ゆーい差」への強い執着が読み取れる．

今日のぽいんと: いつも「検定って何?」再検討

1. **検定って何なの**

(Parametric) Bootstrap 法による検定を使えるようになろう!

帰無仮説, 危険率, 棄却域, 有意水準 ……

2. **検定はそんなにエラいのか ?!**

二種類の過誤と非対称性

3. **検定つかうときの注意点**

標本数に注意!

より良い推定結果を選ぶ方法として**モデル選択**が最近はよく使われるようになってきた

(今回は説明できませんでした)