2004.11.29

生物多様性論 I: 「統計モデルの基礎」

全部で 2 回講義の 2

「検定」の使われかたを観察しよう

— 「検定」はそんなにエラいのか? —

http://hosho.ees.hokudai.ac.jp/~kubo/stat/2004/

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**: 2004.11.24 (**水**)
  統計モデリングと推定を重視してみる
  — 理解できる統計学めざして
- **第 2 回**: 2004.11.29 (**月**)
  「検定」の使われかたを観察しよう
  — 「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

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．

以上のやりとりから，
浦井君の「ゆーい差」への強い執着が読み取れる．

今日のハナシ: 検定とは何か? エライのか?

1. **検定って何なの**

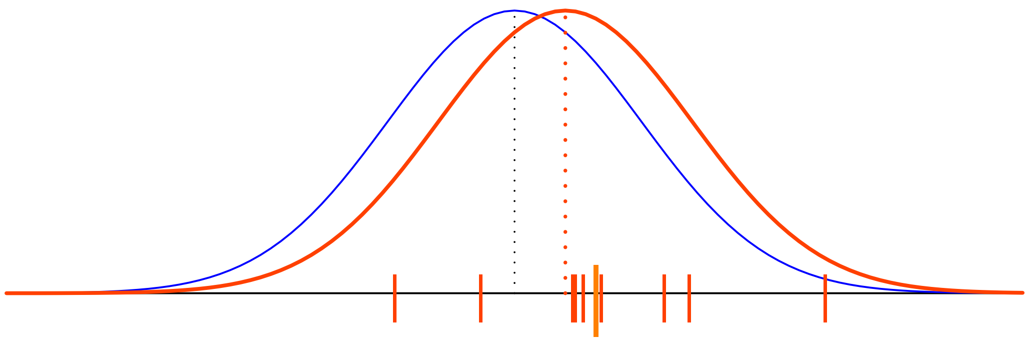
   コイン投げの具体例で考える

2. **検定という操作を統計学用語で言い直す**

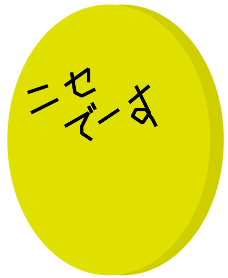
   コイン投げの例を難しく表現してみる

3. **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



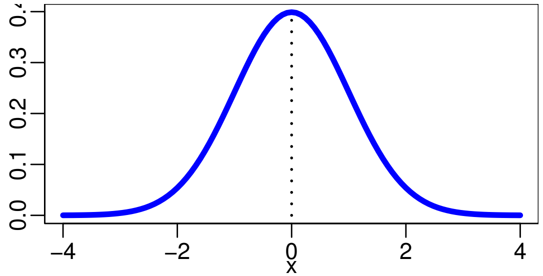
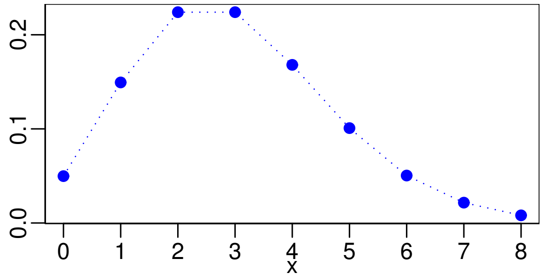
統計学はニセ金見破りに使えるのか?



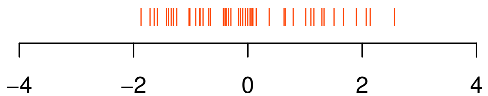
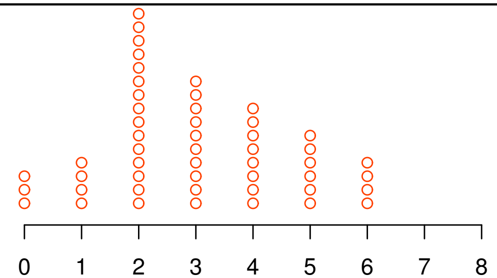
1. 検定って何なの

コイン投げの具体例で考える

(前回の復習) 乱数 (標本) と推定の関係

(よびかた)	[連続確率密度分布]	[離散確率密度分布]
• モデル • 確率分布 • 母集団		

サンプリング ⇓ ⇑ (パラメーター) 推定

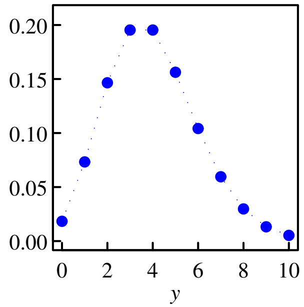
• データ • 乱数 • 標本集団		

- 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 → 現象を簡単な数式で書ける
- 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っていく (今回のハナシ)

(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

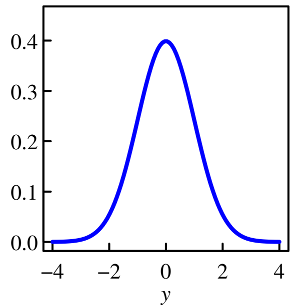
5 4 3 2 4 2 4 1 7 1 →



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
のパラメーターを決めてやる技法

(分布がわかってないと推定できない)

さて今日の問題: 偽造コインのオモテ・ウラ

コイン投げやったときの「オモテの出る確率」でコインの真贋がわかるとする

- **本物コイン**— **オモテが出る確率はぴったり** 50% ($q = 0.5$)
- **偽物コイン**— **オモテが出る確率は** 50% **ではない** ($q \neq 0.5$)

ここに「いかにもニセっぽい」コインあり，20 回コイン投げやったときにオモテが出た回数は 6 回だった．これはニセものと言えるか?

[**考えの足りない回答例**]

- **オモテが出た確率は** $\frac{6}{20} = 0.3 \neq 0.5$，0.2 **もずれてる，つまりニセものぢゃん!**
  - **ホントか?** (**これはあとで検討**)
- **なンか統計ソフトに入れて「ゆーい差」の出る検定やりゃあいいんでしょ!**
  - **あんたは「浦井くん」か? どういう「検定」やるつもり?**

検討の出発点: もしこれが本物だったら?

実際のところ「いかにもニセっぽい」けど，それはいったんおいて

- **本物だったらオモテがでる確率は** 0.5 ($q = 0.5$)
- 20 **回なげたら何回オモテがでるか**?
- **いつも** 10 **回オモテがでるわけないよな**
- 9 **回とか** 11 **回ってのもけっこうありそう**…… ?

統計学でわからないことがあったら「実験」せよ!

今回も活躍する統計ソフトウェア R

`http://www.r-project.org/`

R でコイン投げをする: $q = 0.5$ を 20 回

- **確率 p でオモテ, $1-p$ でウラとなる確率論的モデルはベルヌーイ試行という**
- **ベルヌーイ試行は R の `rbinom()` 関数でシミュレイトできる (本物コイン のシミュレイション — 乱数の生成)**
- **オモテを 1, ウラを 0 であらわす**

```
> rbinom(n = 20, size = 1, prob = 0.5)
[1] 0 1 1 1 0 0 1 1 0 0 1 1 1 0 1 1 0 0 0 0
> rbinom(n = 20, size = 1, prob = 0.5)
[1] 1 0 1 1 1 1 1 0 0 0 0 0 1 1 1 0 0 1 1 1
> rbinom(n = 20, size = 1, prob = 0.5)
[1] 0 0 1 1 0 0 0 0 0 0 1 1 1 1 0 0 1 1 0 0
```

本物コイン投げシミュレイションを何度もやる

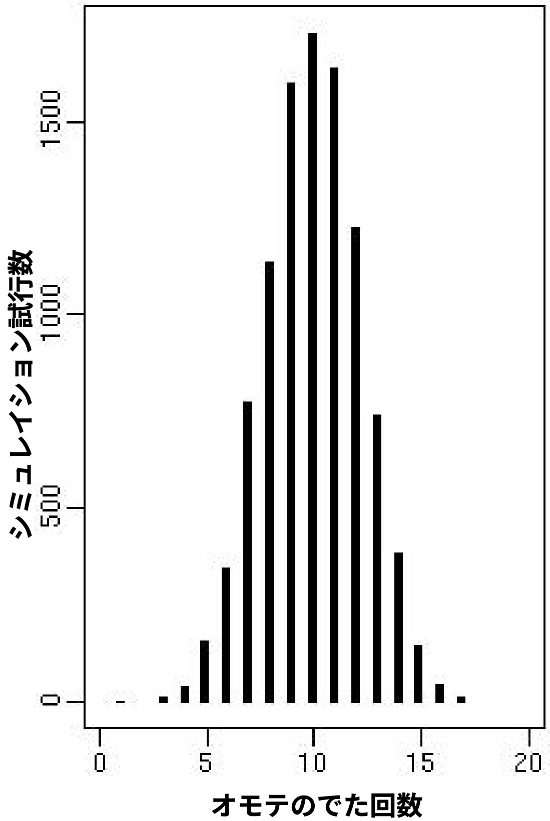
- sum() 関数つかうと 20 回試行中のオモテの回数わかる

```
> sum(rbinom(n = 20, size = 1, prob = 0.5))
[1] 10
> sum(rbinom(n = 20, size = 1, prob = 0.5))
[1] 12
> sum(rbinom(n = 20, size = 1, prob = 0.5))
[1] 9
```

- これをなんどもやる → 100 回やる → 1000 回やる → ……

```
> sapply(1:100, function(i) sum(rbinom(20, 1, 0.5)))
  [1] 10  6  9 11  9 10 11 11  7  9  7  9  9 12 11 12  9  6 12 12  9 10 11  7 13
 [26] 11 12  9  9 10 11  8 12  8  7 10 12  9 11  8 12  9 10 13 13  6  9 11  8 11
 [51]  8  9 10 13 11  9 14 12 11 14  8 13 11 11  6  6  8 10  8 11  9 11  9  6  7
 [76]  9  9  8  9 10  9 11  8 10 10  8  9 11  7  7 10 11  9 12 11  8 10  9  7 14
```

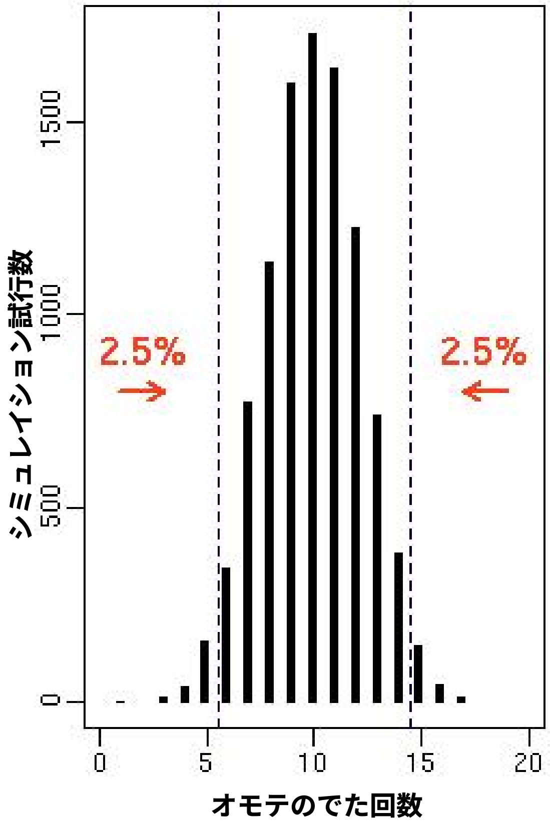
本物コイン投げ: 20 回 × 10000 回やると



- (左の図) 横軸は 20 回コイン投げのうち，オモテの出た回数 x $(0 - 20)$; 縦軸は 10000 回試行やったときに x オモテになった回数

```
n <- 10000 # シミュレイション回数
p <- 0.5   # オモテのでる回数
n.sample <- 20 # 何回コイン投げるか
range.x.char <- as.character(0:n.sample)
trials <- sapply(
    1:n,
    function(i) sum(rbinom(n.sample, 1, prob = p))
) # これは前のスライドでやった繰り返しワザ
count <- sapply(
    range.x.char,
    function(k) sum(trials == as.numeric(k))
) # 「オモテ x 回」が何度あったか数える
```

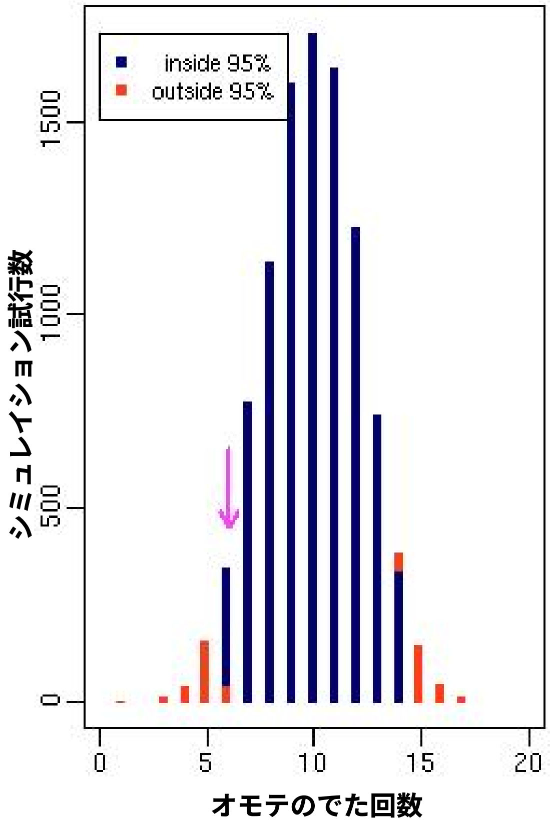
本物コインでは「めったに出ない」領域さがす

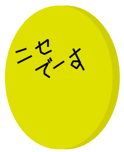


- **けっこうばらつくことがわかった**
- $10000\cdots000$ **回ぐらいやればオモテ** 0 **回とか** 20 **回なんかも出るかも**?
- **妥協**: **本物コインで「**20 **回なげたときに** x_1 **回以下および** x_2 **回以上は めったにでない」という領域をさがす**
- **「めったにない** 5%**」あたりにしよう**
- **下** 2.5% **と上** 2.5% → **下** 250 **試行と 上** 250 **試行**
- **前スライドの** `count` **をチェックしてみると**:

```
> count # 左のグラフと対応
   0    1    2    3    4    5    6    7    8    9   10
   0    0    1    9   52  134  367  728 1205 1628 1768
  11   12   13   14   15   16   17   18   19   20
1643 1154  683  407  166   46    7    2    0    0
```

さて「20 回中の 6 回がオモテ」は……?





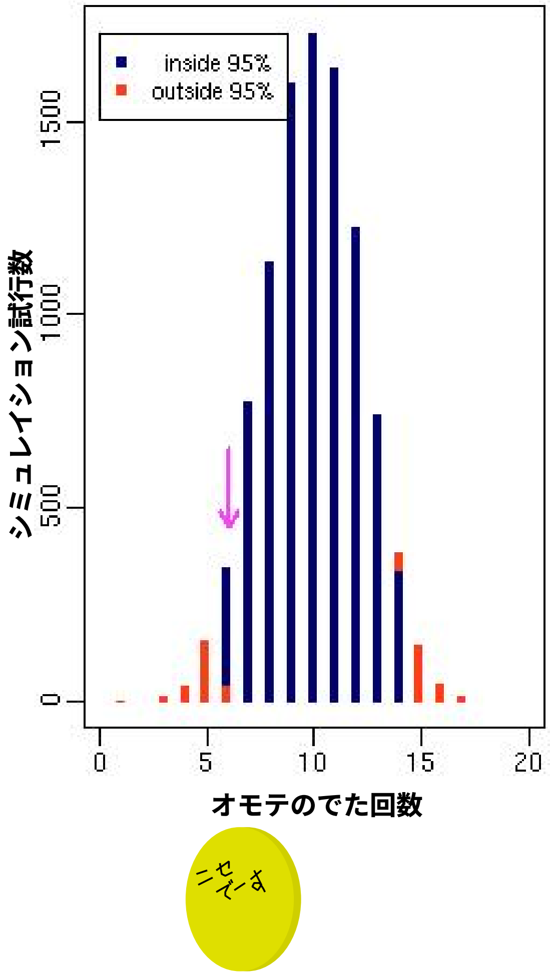
- **下から数えてみると ……**

x	(**累積**)	**確率**
0, 1	0 **試行**	0.000
2, 3	10 **試行**	0.001
4	62 **試行**	0.006
5	196 **試行**	0.020
6	563 **試行**	0.056

- **上から数えてみると ……**

x	(**累積**)	**確率**
20, 19	0 **試行**	0.000
18, 17	9 **試行**	0.001
16	51 **試行**	0.005
15	217 **試行**	0.021
14	624 **試行**	0.062

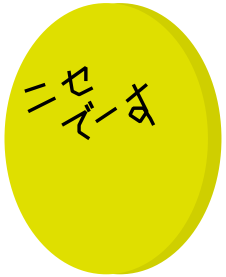
検定の結論: これでは「ニセ金」と断定できん



- きわどい領域をみる:
  1. x が 5 以下または 15 以上: 確率　4.1%
  2. x が 6 以下または 14 以上: 確率 11.8%

  「めったにない 5%」あたりを探していたので 1. にするか
- すると「20 回中 6 回がオモテ」は「よくある 95.9% 領域」に含まれる
- したがって，これは 偽物コイン とは言えない
- とはいえ,「本物コイン」と言えるわけでもなし……すっきりしない解決
- 「オモテ 5 回」だったら断定できたか?

統計学本とかが読めるようになるために



2. 検定という操作を統計学用語で言い直す

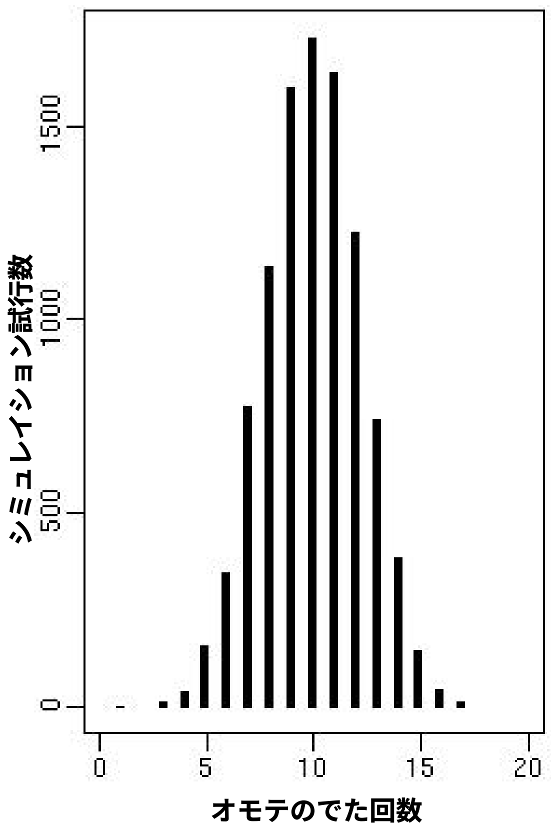
コイン投げの例を難しく表現してみる

検討の出発点: もしこれが本物だったら?

実際のところ「いかにもニセっぽい」けど，
それはいったんおいて

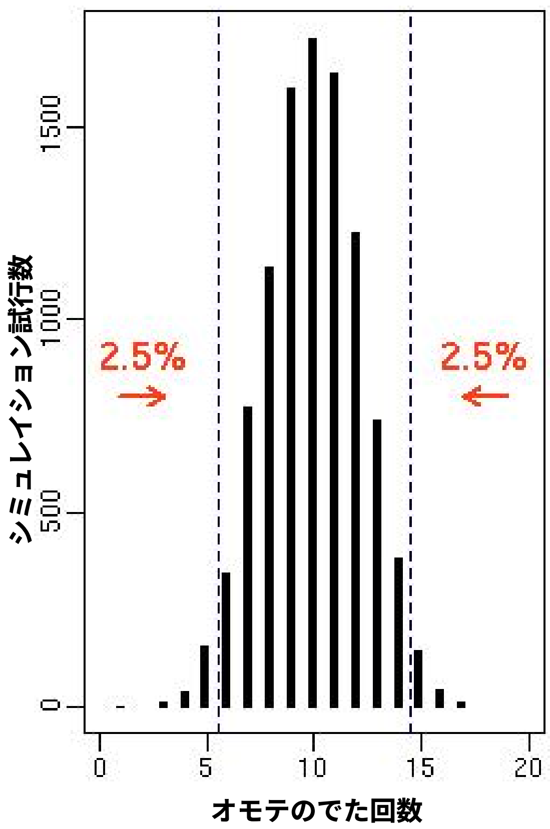
- 本物コインはオモテがでる確率 $q = 0.5$ であるとわかっている
- いま手元にある「いかにもニセっぽい」コイン C はオモテが出る確率が本物とは異なるゆえに，偽物であると示したい
- 検定のための帰無仮説 (あまり成立してそうにない仮説 / 捨てたい仮説) を以下のように定める:
  - 「いかにもニセっぽい」コイン C においてもオモテがでる確率が 0.5 である
- この 帰無仮説を (統計学的) 検定によって棄却したい
- 「20 回コイン投げしたときオモテのでる回数」が検定統計量

本物コイン投げ: 20 回 $\times$ 10000 回やると



- 「いかにもニセっぽい」コイン C を 20 回投げる実験をする前に **帰無仮説** (「コイン C のオモテがでる確率は 0.5」) のもとで得られる**母集団**について考える
- オモテがでる確率 $q = 0.5$ のもとで,「20 回中オモテが x 回出現する」(つまり**検定統計量**の) 確率分布をモンテカルロシミュレイションによって生成した
- 母集団から**検定統計量の確率分布**を生成するモンテカルロシミュレイションの試行回数は 10000 回とした

本物コインでは「めったに出ない」領域さがす

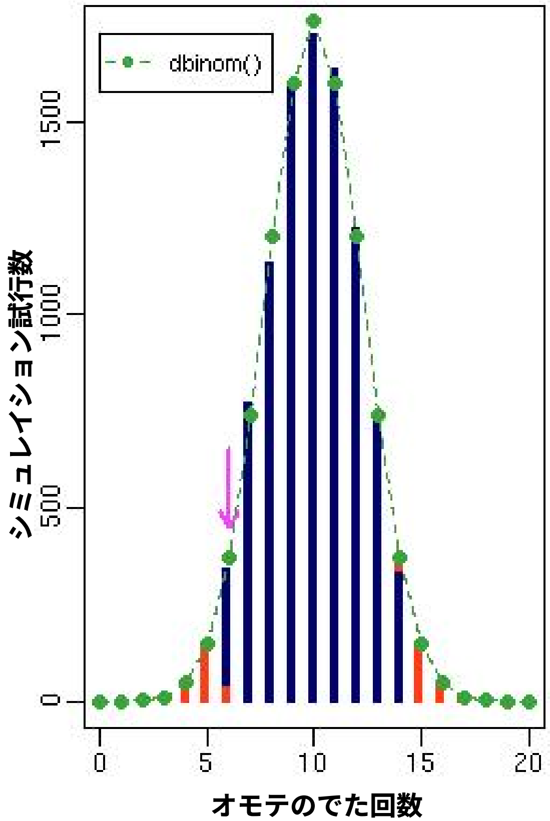


- このようにして得られた確率分布 (母集団) において**棄却域**を構成したい
- 左の図から計算すると，
  1. x が 5 以下または 15 以上: 確率 **4.1%**
  2. x が 6 以下または 14 以上: 確率 11.8%

  となることがわかった
- そこで**危険率**(あるいは **有意水準**) 4.1% となるように棄却域を構成する
  → ニセっぽいコイン C を 20 回投げたときにオモテ出る回数 X_C が「5 以下または 15 以上」が**棄却域**

危険率は好きなように決めてよい (5% でなくても構わない)

[別解] 二項分布を利用した計算

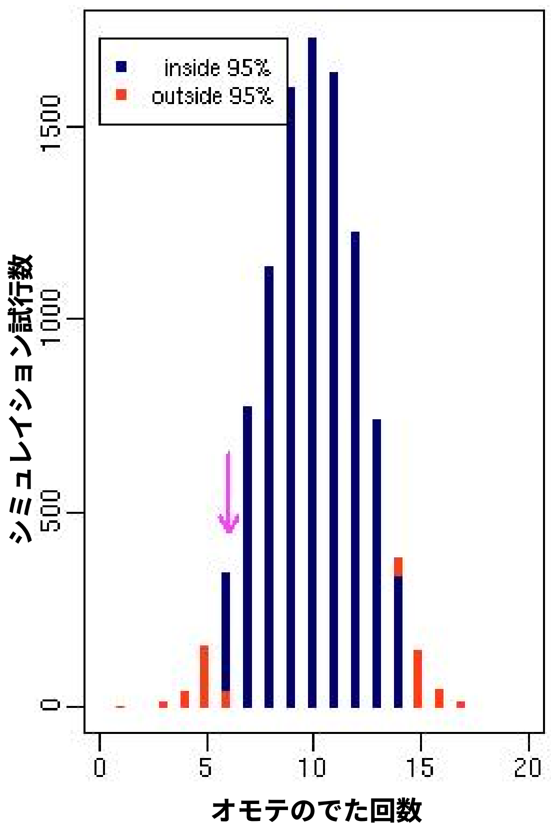


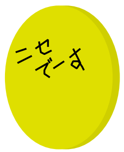
- この問題の場合，モンテカルロシミュレイションをやらなくても，母集団となる確率分布や棄却域を計算できる
- オモテでる確率 $q = 0.5$ のコインを 20 回なげたときに x 回オモテがでる確率は

$$\Pr(X = x) = \frac{20!}{x!(20-x)!} q^x (1-q)^{20-x}$$

- R だと `dbinom(x, 20, 0.5)`
- この二項分布による確率分布 (検定統計量 の分布) の計算は,「モンテカルロシミュレイションを無限回やった」場合と同じ結果をだす

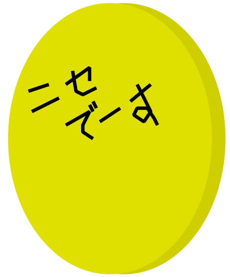
検定の結論: これでは「ニセ金」と断定できん





- **実験**: **ここまで準備してから,「いかにもニセっぽい」コイン** C **を** 20 **回投げる**
- **実験結果**: 20 **回中** 6 **回がオモテであった** ($X_C = 6$)
- **これは事前に定めた棄却域「**X_C **が** 5 **以下または** 15 **以上」に含まれない**
  - (**別の表現**) $X_C = 6$ **となる** p **値** (**最小の有意水準**) **は** 0.118 **である** ($p = 0.118$)
- **したがって 帰無仮説「コイン** C **においてオモテのでる確率は** 0.5 **である」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** $X_C = 5$ **だったら**)
  帰無仮説は危険率 4.1% ($p = 0.041$) **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**

道具を「理解」して使うのが自然科学の作法

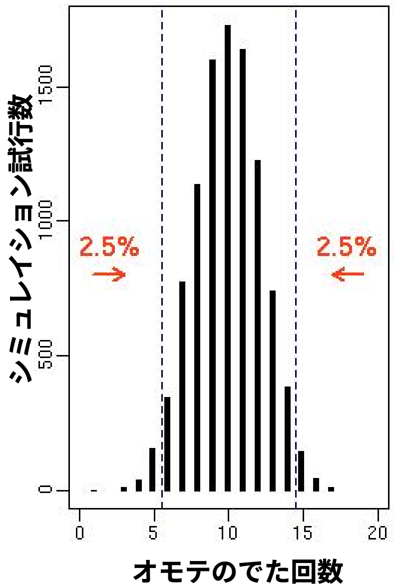


3. 検定はそんなにエラいのか

そのあたり，検討してみましょう

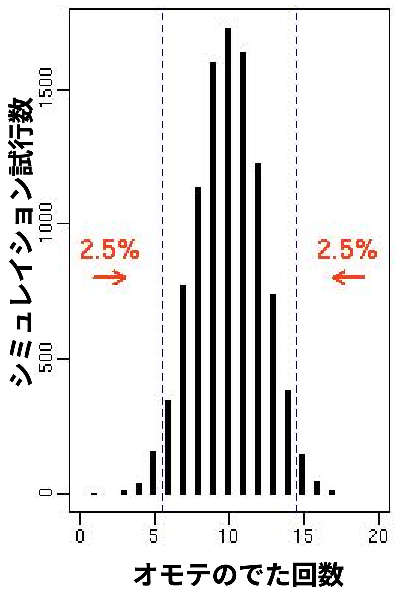
検定についてまわる二つの過誤

↓帰無仮説は	X_C が 棄却域内 (帰無仮説棄却)	X_C が 棄却域外 (棄却できない)
正しい	第一種の過誤	OK
正しくない	OK	第二種の過誤



- 第一種の過誤
  - 真実: 「ニセものっぽいけど本物」
  - 検定結果: 「ニセものと断定する」
- 第二種の過誤
  - 真実: 「正真正銘のニセもの」
  - 検定結果: 「ニセものと断定できない」

検定の非対称性



- **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「ニセものっぽいけど本物」の可能性 (危険率, p 値) を設定・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**
  - **間接的には考慮する場合がある**
  - **直接的には 検定力(あるいは検出力) 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

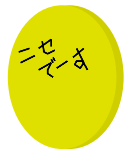
– **帰無仮説が棄却できない場合 → 何も言えない**

なぜこのような「非対称性」があるのか?

検定にはなぜ非対称性があるのか?

(検定の特徴) 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するのか

答: いま調べてるコインが「ニセっぽい」から



— **帰無仮説さえつぶせばいい** (問題が簡単になる)

- 「こんなにニセっぽいんだから，帰無仮説 (「オモテの出る確率は 0.5」) なんてのは棄却されて当たりまえだ」と見なされてる
- 科学の実験での例:
  - 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)

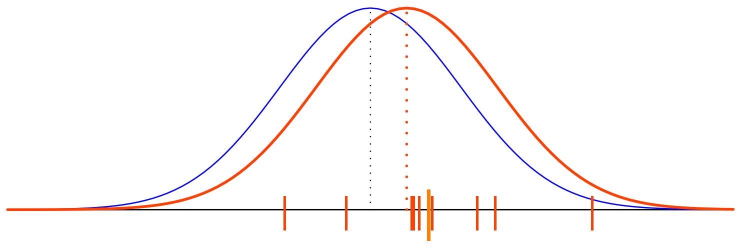
検定を使ううえでの諸注意

- **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)**
- **検定の解釈について— 再びコイン投げ問題を考えよう**
  20 回投げて 6 枚オモテvs 200 回投げて 60 回オモテ

```
> pbinom(6, 20, 0.5) * 2   # オモテ確率 0.3
[1] 0.11532
> pbinom(60, 200, 0.5) * 2 # オモテ確率 0.3
[1] 1.5071e-08
> pbinom(85, 200, 0.5) * 2 # オモテ確率 0.425, ついでながら
[1] 0.040037
```

- **この例からわかることは……**
  - **有意差でるかどうかは 標本数に依存する**
  - 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**

今日のまとめ: 「検定」使用は理解が必要



1. **検定って何なの**

   コイン投げのような具体例をいつでも構成できるように

2. **検定を統計学用語で言い直す**

   帰無仮説，危険率，棄却域，有意水準 ……

3. **検定はそんなにエラいのか ?!**

   あなたの研究に **検定は必要か?**

   より良い推定結果を選ぶ方法としてモデル選択もある
   (今回は説明できませんでした)