

## 動物実験統計学の実際

# 実験動物統計学におけるエビデンスの 蓄積・解析・プレゼンテーション

東北大学大学院医学系研究科循環器内科学分野

宮田 敏

E-mail: miyata@cardio.med.tohoku.ac.jp

### Agenda

1. データの準備
2. データの提示
3. 推定、信頼区間、仮説検定
4. (時間があれば) 経時データの解析

## 1. データの準備

### i. データの形は長方形

- 第一行目に**変数名**。全角文字は**避ける**方が無難。
- グラフ、解析結果などを**張り付け**ない。別ファイルで保存。
- データの形は、**長方形**になるはず。

systemID	hospitalID	sex	age	height	bodyweight	
4	1185645		1	64	173	75.4
11	3329388		1	69	164	72
12	4022624		1	78	155.2	47.2
14	4402536		1	83	159.1	60
22	4862866		2	73	147.6	40.5

## 1. データの準備

### ii. 元データは絶対に改変しない。

- 解析の過程で、変数を変換したり、新しい変数を定義することがある。
- 新しく作ったデータを、元データに**上書き**しない。
- データを改変したら、新しいファイル名で保存。
- 元データを改変すると、元データが何であるか分からなくなる。元データが分からなくなれば、**意図せざるデータのねつ造**まであと一歩。

## 1. データの準備

### iii. 解析記録の保存。

- 実験をすれば、実験ノートに記録するのは常識。統計解析の記録を残すのも、それと同じ。
- 元データと解析の記録を見れば、第三者が解析を再現できる程度の記録が必要。
  - 解析の再現性 (Traceability, 追跡可能性)
  - 備忘録 「三日後の自分は遠い親戚。一週間後の自分は赤の他人」
  - 出来れば、プログラムを書いて解析する。

## 1. データの準備

### iv. データ入手時にすべきこと：入力ミス、異常値の発見

表計算ソフトのフィルター機能が便利

- データの範囲： 本来正の値をとるはずが、負の値をとる。体重35kgが3.5kgになる、等。
- 全角文字と半角文字の混在： “w” と “w” など。
- 質的変数の数字表記： 男性→1, 女性→2など。男性→M, 女性→F のように書き直す。
- 異常な値の検出： “3.14” と “3,14” など。
- 欠測値の数： 欠測値の数が想定より多い場合、データが正常に認識されていないことがある。

## 2. データの提示 —Table 1のまとめ方—

得られたデータは、サンプルの群ごとに分類し**基本統計**

### i. 連続数（実数）の提示

- Mean  $\pm$  SD (Standard Deviation)
- Mean  $\pm$  SE (Standard Error, SD/ $\sqrt{N}$ )
- Median (IQR, Inter quartile range)

第一選択は、**Mean  $\pm$  SD**（データ全体の分布の比較）または、**Mean  $\pm$  SE**（平均の比較）

Mean  $\pm$  SDの上限、下限を実際に計算して、その変数が通常とる範囲を外れるときは、**Median (IQR)**を選択。

## 2. データの提示 —Table 1のまとめ方—

### ii. 離散数（カウント）の提示

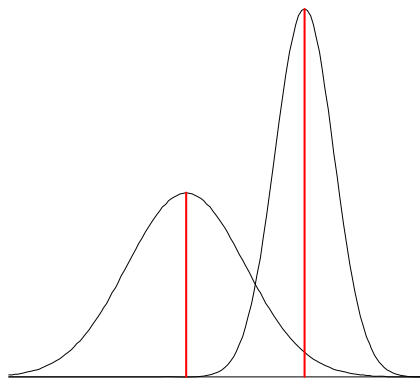
- **度数**（個数、frequency）と**パーセント**を両方提示する。

度数のみ、あるいはパーセントのみを提示した論文を見かけるが、お勧めしない。必ず両方出す。

比較のための検定（必ず**p値**を記載する）

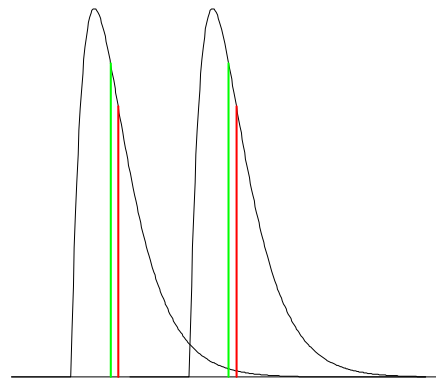
- 連続数：Mean  $\pm$  SE  $\Rightarrow$  **Welch' s t-test**（不等分散）  
Median (IQR)  $\Rightarrow$  **Mann-Whitney test**,
- 離散数：**Fisher' s exact test**（フィッシャーの直接法） or  **$\chi^2$ 検定**

## T検定に適した比較



- 二群とも正規分布に従う
- 分散は異なってもよい

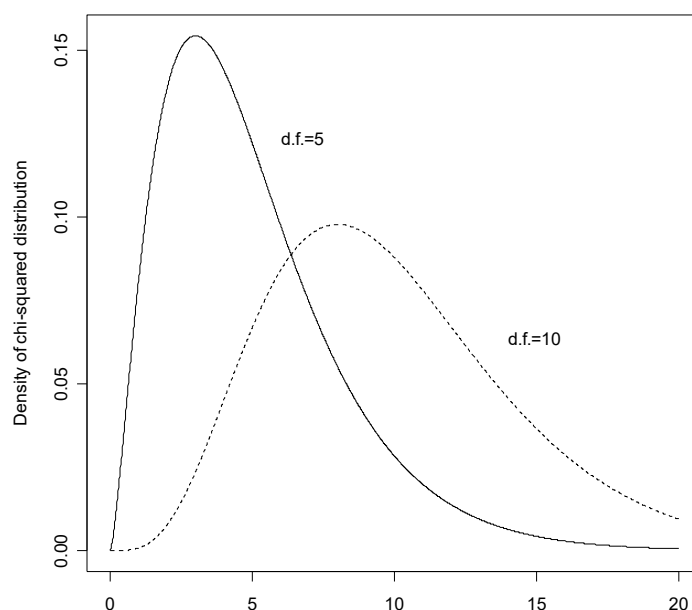
## MW検定に適した比較



- 正規分布に従わなくてもよい
- 分布の形状は同じ. 分散も同じ.

データの分布が正規分布に従わず、分散も等しくない場合、  
取りあえず元データを**対数変換**するなどして、**等分散**に近づける。  
それでもだめなら、専門家にご相談ください。

このような分布は、ノンパラの検定には適さない



データ解析では、様々なパラメーターが推定および検定される。

- 期待値  $\mu = E(X)$
- 割合（イベントの発生確率）
- 回帰係数（傾き）  $Y_i = \beta_0 + \beta_1 x_{1i} + \Lambda + \beta_k x_{ki} + \varepsilon_i$
- オッズ比  $e^\beta$   $\log(p_i/(1-p_i)) = \beta_0 + \beta_1 x_{1i} + \Lambda + \beta_k x_{ki}$   
ロジスティック回帰
- ハザード比  $e^\beta$   $h(t | x_i) = h_0(t) \exp\{\beta_0 + \beta_1 x_{1i} + \Lambda + \beta_k x_{ki}\}$   
Cox比例ハザードモデル

## 1. 推定

パラメーターは、サンプルから計算した推定量 (estimator) を用いて推定される。

- 期待値  $\Rightarrow$  標本平均  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$
- 割合  $\Rightarrow \hat{p} = \frac{X}{n}$   $X$ : イベント数,  $n$ : 試行回数

その他のモデルの推定量も、統計ソフトで計算できる。

## 1. 推定（つづき）

推定量の観測値=推定値 (estimate) をもって、パラメターの値を推定する.

推定量はサンプルから計算されるが、サンプルには観測誤差が含まれる.

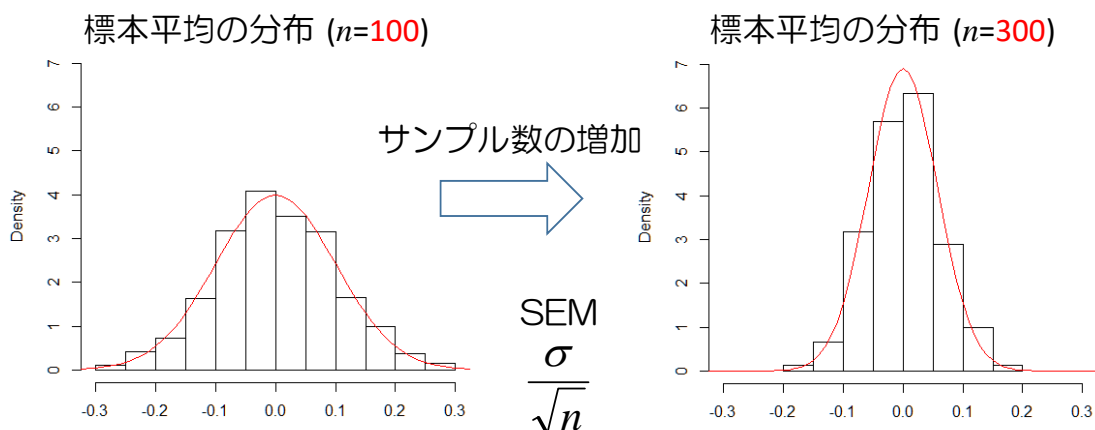
⇒ 推定量にも、観測誤差=分散がある.

⇒ サンプルングのたびに、推定値にはばらつきがある.

## 1. 推定（つづき）

推定量の観測値=推定値からは、推定の精確さは評価できない.

サンプル数が大きくなると、推定量の分散あるいは標準誤差 (standard error) は小さくなる.



## 2. 信頼区間

推定量を標準誤差で割ったものは、標準正規分布（あるいは $t$ 分布）などに従う。

期待値の信頼区間  $\left( \bar{X} - t_{\alpha/2, n-1} \frac{s}{\sqrt{n}}, \bar{X} + t_{\alpha/2, n-1} \frac{s}{\sqrt{n}} \right)$

サンプル数 $n$ が大きくなると、信頼区間の幅は小さくなる。

推定量と信頼区間の組み合わせで、観測誤差も含めたパラメーターの量的な推測が可能になる。

## 3. 仮説検定

帰無仮説 (Null hypothesis) :  $H_0$  母集団に関する命題で、検定の当初、真であると仮定されるもの。

対立仮説 (Alternative hypothesis) :  $H_1$  帰無仮説  $H_0$  と対立する仮説。

Type I error :  $H_0$  が真であるとき  $H_0$  を棄却する過誤

Type II error :  $H_0$  が偽であるとき  $H_0$  を棄却しない過誤

有意水準 (Significance level)  $\alpha$  : 事前に決められた定数で、検定は  $P(\text{type I error}) = \alpha$  となるように定められる



### 3. 仮説検定（つづき）

仮説検定は，帰無仮説を棄却するか否かを結論するだけ。仮説検定で，**効果 (effect)** の大きさを評価することは出来ない。

単純に，サンプル数を増やせば帰無仮説を棄却しやすく，p値が小さくなる。

帰無仮説が棄却されたからと言って，**効果量 (effect size)** が大きいことを意味しない。

**検出力 (Power) (1-β):**  $H_1$ が真であるとき， $H_0$ を棄却し $H_1$ を採択する確率。

例：A群とB群の体重の期待値を比較する

$\mu_A, \mu_B$  : A群とB群の体重の期待値

$\bar{X}_A, \bar{X}_B$  : A群とB群の体重の標本平均

$\bar{X}_A = 60.2, \bar{X}_B = 60.0$  であるとする。

仮説1 :  $H_0 : \mu_A = \mu_B$  vs.  $H_1 : \mu_A > \mu_B$  **p<0.01**

仮説2 :  $H_0 : \mu_A - \mu_B = 0.1$  vs.  $H_1 : \mu_A - \mu_B > 0.1$  **p<0.01**

どちらの帰無仮説も棄却された。（十分あり得る）

- どちらの仮説が正しい？
- 0.1kgの差に，意味がある？

## 統計的有意性とP値に関するASA声明

1. P値はデータと特定の統計モデルが矛盾する程度をしめす指標のひとつである。
2. P値は、調べている仮説が正しい確率や、データが偶然のみでえられた確率を測るものではない。
3. 科学的な結論や、ビジネス、政策における決定は、P値がある値（訳注: 有意水準）を超えたかどうかによりのみ基づくべきではない。
5. P値や統計的有意性は、効果の大きさや結果の重要性を意味しない。

参照：統計的有意性とP値に関するASA声明

<http://www.biometrics.gr.jp/news/all/ASA.pdf>

2018/5/17

動物実験統計学の実際

19

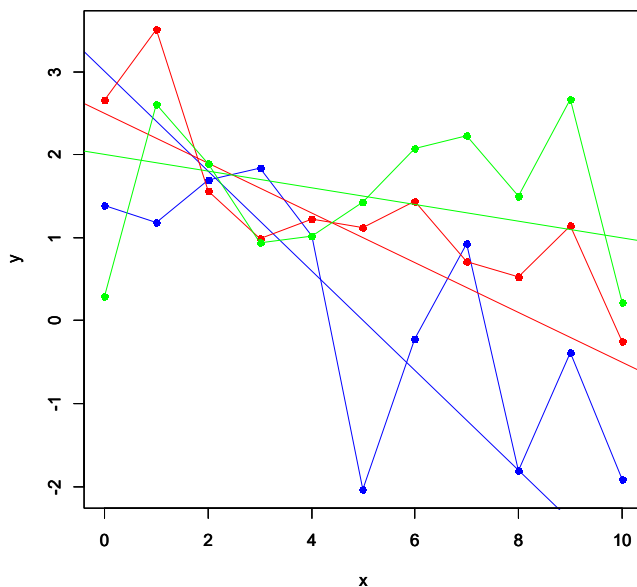
# 正規混合効果モデル

反復測定データに対する分散分析モデルは

1. 各個体を識別するため**多数のパラメータ**が必要  
⇒ 推定の「切れ味」が下がる。
2. 「時間」が、時点の主効果の水準として扱われ、**実数値としての意味を持たない**。

と言う欠点を持つ。

それぞれの個体の持つ個体差を，群ごとの平均的トレンドからのランダムな（=変量効果）乖離としてモデル化する ⇒ **正規線型混合効果モデル**



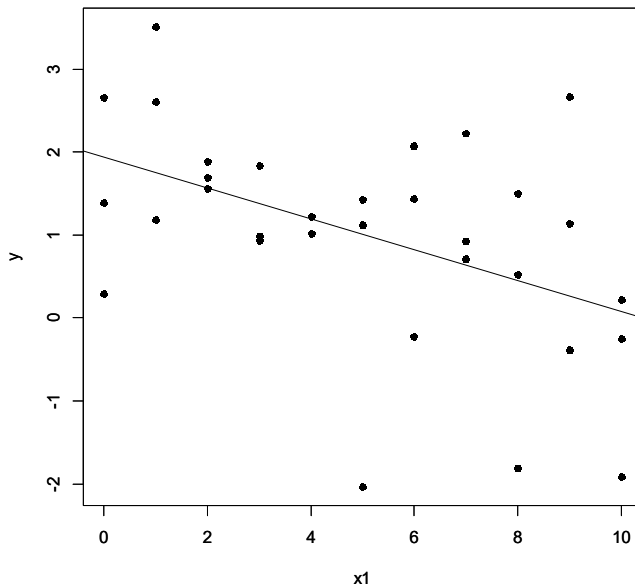
いま仮に，以下Model 1, 2, 3のデータがあったとする。

$$\text{Model1: } y = 3 - 0.6x + \varepsilon$$

$$\text{Model2: } y = 2.5 - 0.3x + \varepsilon$$

$$\text{Model3: } y = 2 - 0.1x + \varepsilon$$

1. それぞれのモデルは，線形で近似できる。
2. それぞれ切片と傾きは異なる。
3. 全体としては，右下がりの傾向がある。



Coefficients:

もし、モデルを無視してデータをすべてまとめてしまうと、

1. 当然のことながら、各モデルの特徴は失われる。
2. この場合、傾きが小さ目に推定される。
3. 全てのモデルを推定するには（切片、傾き）× 3 = 6個のパラメーターが必要。

## 多くのパラメーターを推定することの問題点

1. 多くのものを一度に推定すると、推定の分散が大きくなる（自由度が小さくなる） ⇒ 有意差が出にくくなる。
2. 全体としての傾向は、推定できない。

$$\text{Model1: } y = 3 - 0.6x + \varepsilon$$

$$\text{Model2: } y = 2.5 - 0.3x + \varepsilon$$

$$\text{Model3: } y = 2 - 0.1x + \varepsilon$$

必要なパラメーター数

= 7

モデル数が増えると、この差は大きくなる



**線形混合効果モデル (linear mixed effect model)**

$$y = \beta_0 + b_0 + (\beta_1 + b_1)x + \varepsilon,$$

$$b_0 \sim N(0, \sigma_{b_0}^2), b_1 \sim N(0, \sigma_{b_1}^2), \varepsilon \sim N(0, \sigma^2),$$

必要なパラメーター数

= 5

```

> item <- c(rep("I", 11), rep("II", 11), rep("III", 11)) モデルのindicator
> Data <- cbind.data.frame(y, x1, item)
>
> library(lme4)
> fit.lme0 <- lmer(y ~ x1 + (1 + x1 | item), data=Data)
>
>
> summary(fit.lme0)
Linear mixed model fit by REML ['lmerMod']
Formula: y ~ x1 + (1 + x1 | item)
Data: Data

REML criterion at convergence: 95.1

Scaled residuals:
   Min       1Q   Median       3Q      Max
-2.61999 -0.55320 -0.09196  0.64546  1.82209

Random effects:
 Groups Name      Variance Std.Dev. Corr
 item   (Intercept) 0.09340  0.3056
        x1         0.02681  0.1637  -0.56
 Residual                0.78417  0.8855
Number of obs: 33, groups: item, 3

Fixed effects:
              Estimate Std. Error t value
(Intercept)   1.9365     0.3381    5.728
x1            -0.1863     0.1064   -1.752

Correlation of Fixed Effects:
      (Intr)
x1 -0.589

```

切片 独立変数 混合効果を設定する変数

## 正規混合効果モデル（例題に戻る）

$$y_{ijk} = \beta_0 + \beta_1 x_i + \beta_2 t_j + \gamma(x_i t_j) + \varepsilon_{ijk}, \varepsilon_{ijk} \sim N(0, \sigma^2)$$

$$x_i = \begin{cases} 1: female \\ 0: male \end{cases}, t_j : time, (x_i t_j) : interaction$$



$$y_{ijk} = (\beta_0 + \beta_1) + (\beta_2 + \gamma)t_j + \varepsilon_{ijk} : female$$

$$y_{ijk} = \beta_0 + 0 + \beta_2 t_j + 0 + \varepsilon_{ijk} : male$$

各個体間の個体差（この場合は切片）を，上の母数効果モデルからのランダムな乖離として表現するため，変量効果  $b_{ij}$  を導入する。

$$y_{ijk} = \beta_0 + \beta_1 x_i + \beta_2 t_j + \gamma(x_i t_j) + b_{ij} + \varepsilon_{ijk}, \varepsilon_{ijk} \sim N(0, \sigma^2), b_{ij} \sim N(0, \sigma_b^2)$$

まず、混合効果を無視した、母数モデルを推定してみる。

$$y_{ijk} = \beta_0 + \beta_1 x_i + \beta_2 t_j + \gamma(x_i t_j) + \varepsilon_{ijk}, \varepsilon_{ijk} \sim N(0, \sigma^2), b_{ij} \sim N(0, \sigma_b^2)$$

$$y_{ijk} = 20.25 + 0.80x_i + 0.96t_j + 0.61(x_i t_j)$$

```
> ### 通常の固定効果モデル ###
>
> Orthodont$age <- as.numeric(Orthodont$age)
> fit.lm <- lm(distance ~ age*Sex, data=Orthodont)
> summary(fit.lm)

Call:
lm(formula = distance ~ age * Sex, data = Orthodont)

Residuals:
    Min       1Q   Median       3Q      Max
-5.6156 -1.3219 -0.1682  1.3299  5.2469

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  20.2500     0.8334   24.297 < 2e-16 ***
age           0.9591     0.3043    3.152  0.00212 **
SexMale      0.7969     1.0827    0.736  0.46337
age:SexMale  0.6097     0.3953    1.542  0.12608
```

時間 age の係数が  $0.96 > 0$  であることから、時間経過と共に distance が増加する。性別の主効果は Male で  $0.78 > 0$  より、男性で大きな傾向があるが有意でない。交互作用は  $0.61 > 0$  より、差が開く傾向があるが有意ではない。

## R (lmer/パッケージ) による混合効果モデルの推定結果

```
> fit.lme1 <- lmer(distance ~ age * Sex + (1 | Subject), data=Orthodont)
> summary(fit.lme1)
Linear mixed model fit by REML t-tests use Satterthwaite approximations
lmerMod]
Formula: distance ~ age * Sex + (1 | Subject)
Data: Orthodont

REML criterion at convergence: 431

Scaled residuals:
    Min       1Q   Median       3Q      Max
-3.5980 -0.4546  0.0158  0.5024  3.6862

Random effects:
 Groups Name Variance Std.Dev.
 Subject (Intercept) 3.299  1.816
 Residual          1.922  1.386
Number of obs: 108, groups: Subject, 27

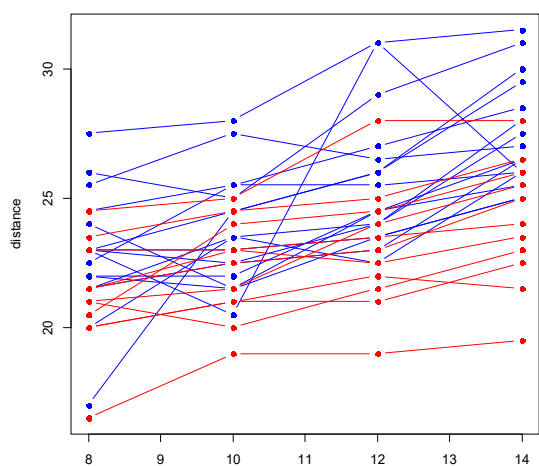
Fixed effects:
            Estimate Std. Error    df t value Pr(>|t|)
(Intercept)  20.2500     0.7496 59.3100  27.013 < 2e-16 ***
age           0.9591     0.1869 79.0000   5.130 2.02e-06 ***
SexMale      0.7969     0.9738 59.3100   0.818  0.4165
age:SexMale  0.6097     0.2428 79.0000   2.511  0.0141 +
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

性別の差は有意では無い登録時(8歳)では、有意差がなかったことを示す。

- ageは有意。係数が正であるから、加齢に伴い、yが増加する。
- 交互作用が有意。交互作用の係数が正の値であるから、男性の場合yの増加率が大きいことを意味する。

係数の推定値は、固定効果モデルと全く同じ。しかし変量効果を導入したことで、ageの主効果と交互作用が**有意**に推定されている。

## 混合効果モデルの続き



```
> confint(fit.lme2)
Computing profile confidence intervals ...
              2.5 %      97.5 %
.sig01      0.86960714  2.7500615
.sig02     -1.00000000  1.0000000
.sig03      0.00000000  0.6771261
.sigma      1.09703160  1.5884953
(Intercept) 18.79934057 21.7006594
age         0.55353088  1.3646509
SexMale    -1.08758687  2.6813369
age:SexMale 0.08282116  1.1364970
```

前項までのモデルは、全被験者共通の傾き、すなわち共通の成長速度を仮定していた。

改めて元データを見ると、被験者によって傾きも違って見えるように見える。

切片と傾きの両方に、変量効果を仮定することも出来る。

この場合は、あまり変わらなかった。

## 混合効果モデルの、その他の使い方

今回は、経時データに対して被験者のbaselineの値に変量効果を入れた。他に、混合効果モデルの使い方には以下のようなものがある。

- 反復測定データあるいは同じ被験者（検体）に異なる介入を行った場合、被験者のbaselineに混合効果を入れる。
- 被説明変数が実数の場合だけでなく、二値の場合（ロジスティック回帰）やその他の分布を持つ場合でも、変量効果を考えられる。

⇒ **Generalized Linear Mixed-effect Model (GLMM)**

## サンプルサイズ的设计

各種の検定において必要なサンプルサイズを设计するには、以下の項目を事前に定める必要がある。

- **有意水準 (significance level)  $\alpha$**  : 帰無仮説が正しいにもかかわらず、誤って帰無仮説を棄却する確率  $\Rightarrow$  対立仮説とは**関係ない**ことに注意。
- **検出力 (power)  $(1-\beta)$**  : 対立仮説が正しいときに、帰無仮説を棄却し対立仮説を採択する確率。
- **Effect size (clinical difference)** : 臨床的に意味のある差。
- **分散** : サンプルの散らばりの大きさ
- **脱落率** : 症例登録ののち、追跡不能になる割合。



# サンプルサイズの設計の手順

1. 当該 RCT に最適な**エンドポイント**の選択
  - 作業仮説の検討に適しているか?
  - 客観的に測定できるか?
2. エンドポイントに期待される **effect size** の見積もり
  - 探索的試験、先行研究から見積もり
  - Cohenの**Effect size index**
  - 可能なサンプル数からの**逆算はタブー**
3. 検定の方法、有意水準、検出力の設定
4. 上記の下で「有意」となる最小の症例数の計算
  - 多すぎても少なすぎてもダメ

## 参考文献：

東北大学大学院医学系研究科循環器内科学分野・  
東北大学病院臨床研究推進センター「医学統計勉強会」

[http://www.cardio.med.tohoku.ac.jp/newmember/medical\\_statistics.html](http://www.cardio.med.tohoku.ac.jp/newmember/medical_statistics.html)