

サンプルサイズ設定



春山康夫

研究連携・支援センター
先端医科学統合研究施設
獨協医科大学

本日の主な内容

- ✓ 臨床研究デザインとサンプルサイズ
- ✓ CONSORT声明とSTROBE声明
- ✓ 帰無仮説と対立仮説
- ✓ P値とは
- ✓ α エラーと β エラー(検出パワー)
- ✓ 両側検定と片側検定
- ✓ サンプルサイズの計算
(手計算・PSソフト・EZRソフト)

臨床研究によく使う研究デザインとサンプルサイズ

- ケース・レポート
- 少数例のケース・シリーズ
- N-of-1 試験

サンプルサイズ
の設定は不要

➤ 観察研究(分析疫学)

ケース・コントロールスタディ
コホート研究

➤ 介入研究

ランダム化比較試験(RCT)
非(準)ランダム化比較試験準実験(non-RCT)
前後比較研究(治験第1相、第2相)
クロスオーバー試験(Cross-over trial)

サンプルサイズ
の設定

CONSORT声明

(Consolidated Standards of Reporting Trials)

2010 checklist of information to included when reporting a randomized trial

7a: Sample sizeはどのように症例数が決まったか。

津谷等. Jpn Pharmacol Ther. 2010.11:939-747

STROBE声明

(The Strengthening the Reporting of Observational Studies in Epidemiology)

10: study sizeがどのように算出されたかを説明する。

科学的証拠はどう取れるか？

例：新しい降圧剤（A剤）に効果があると証明したい



高血圧症患者に対して新しい降圧剤（A剤）を投与し、
従来の降圧剤（B剤）と比較する研究を企画する



下記のAとBは、どのアプローチを取りますか。

A:従来のB剤により新薬に効果があることを示す証拠を挙げる

対立仮説 (H_1)

B:従来のB剤と新薬に効果ないことに反対の証拠を挙げる

帰無仮説 (H_0)

ヘンペルのカラス説

全てのカラスは黒い
(対立仮説)



対立仮説は反証されやすい



Aカラス



Bカラス

...



全てのカラス



1羽のカラス

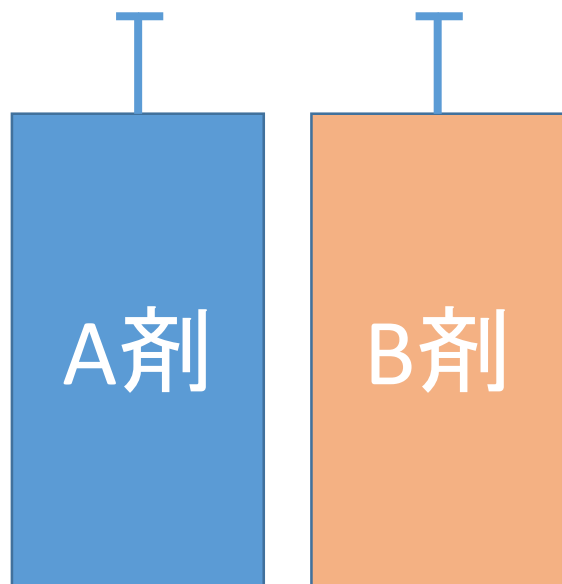
全てのカラスは黒くない
(帰無仮説)

理論的推論

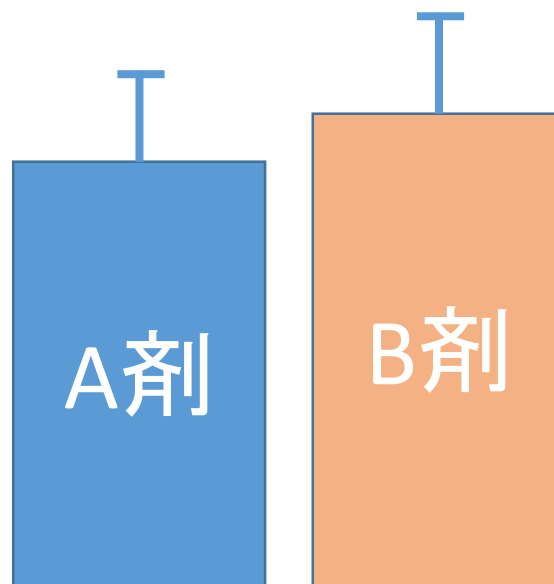
- カラスは黒くない仮説ですが、しかし、見かけたカラスは黒い
- つまり、実際に、黒いカラスの割合が多い事実
- たまに、黒くないカラスは存在するが...
- 黒くないカラスの確率は非常に低い(P 値 <0.05 或いは <0.001)
- そうすると、カラスは黒くないという仮説を捨てて、やはりカラスは黒いと言える。
- これは帰無仮説(カラスは黒くない)を棄却、カラスは黒いと判定できる推論である。

臨床研究

例：高血圧症患者に対して新しい降圧剤（A剤）を投与し、従来の降圧剤（B剤）と比較したい。

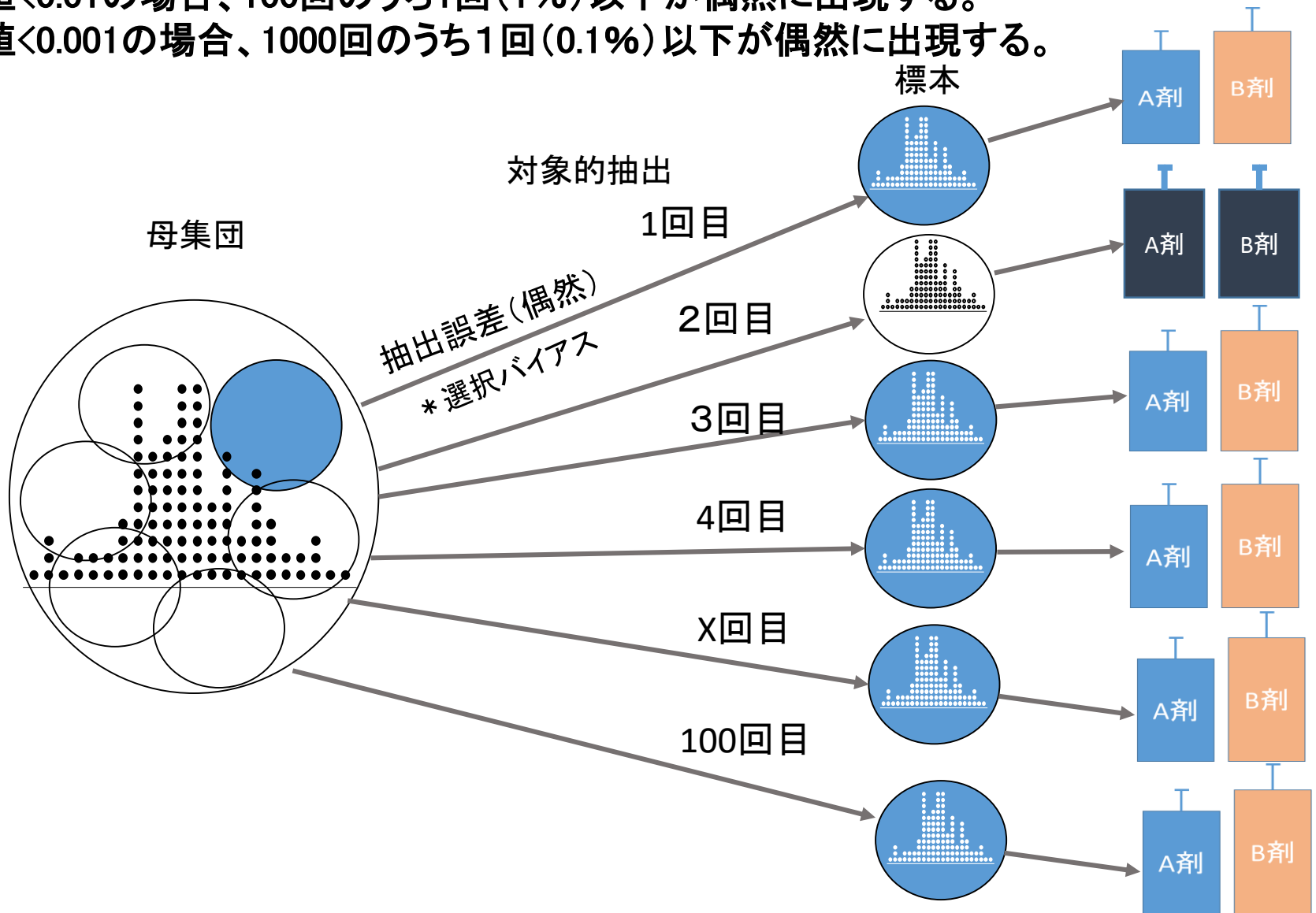


帰無仮説



実際に測定したデータ

P値とは、群間の差がないと仮定した場合、観察された差が偶然に出現する確率
P値<0.05の意味、100回のうち5回(5%)以下が偶然に出現する。
P値<0.01の場合、100回のうち1回(1%)以下が偶然に出現する。
P値<0.001の場合、1000回のうち1回(0.1%)以下が偶然に出現する。



米国統計学会American Statistical Association (2016.3)の声明

(1) P-values can indicate how incompatible the data are with a specified statistical model.

P値はデータが特定の統計モデルにどの程度適合しないかを示すことができる。

(2) P-values do not measure the probability that the studied hypothesis is true, or the probability that the data were produced by random chance alone.

P値は、研究仮説が真である確率あるいはデータがランダムな偶然によってのみデータが得られる確率を示さない。

(3) Scientific conclusions and business or policy decisions should not be based only on whether a p-value passes a specific threshold.

科学的な結論とビジネスや政策の意思決定が、P値が特定の閾値を超えるかどうかのみに基づいてなされるべきではない。

(4) Proper inference requires full reporting and transparency.

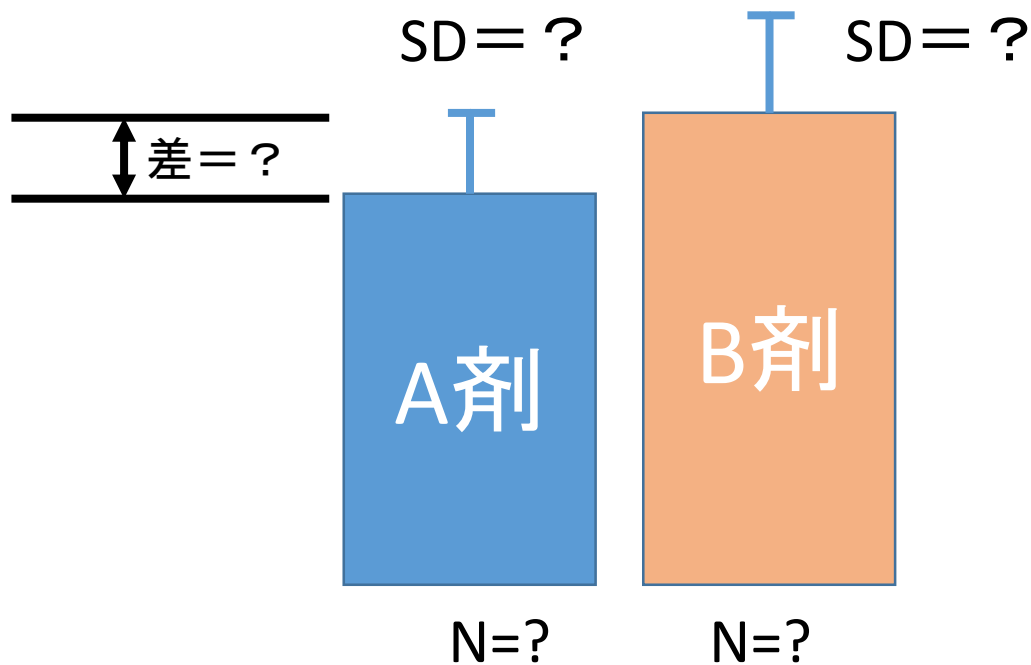
適切な推論は、完全な報告と透明性を必要とする。

(5) A p-value, or statistical significance, does not measure the size of an effect or the importance of a result.

P値もしくは統計学的有意性は、効果の大きさや結果の重要性を測るものではない。

(6) By itself, a p-value does not provide a good measure of evidence regarding a model or hypothesis.

P値そのもののみにモデルや仮説に関するエビデンスの良好な尺度は得られない。



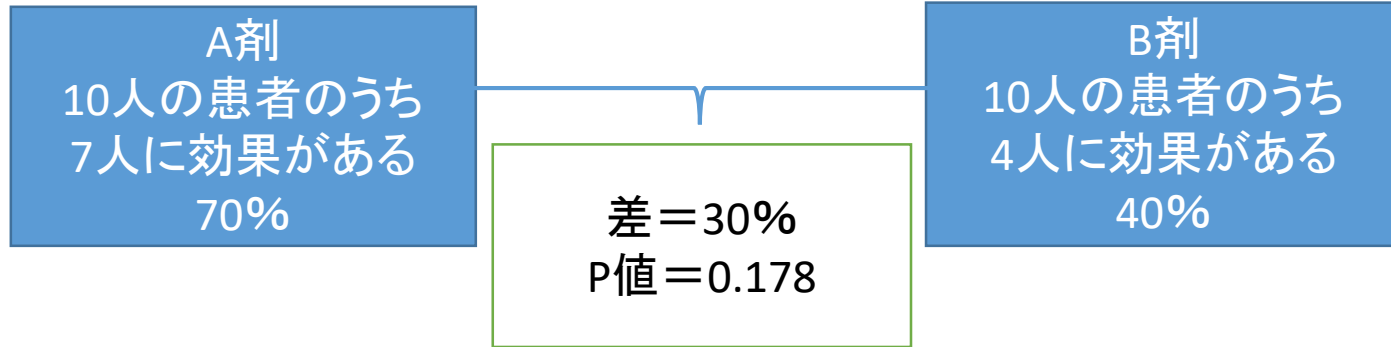
実際に測定したデータ

有意水準(危険率)
P値(確率) = ?

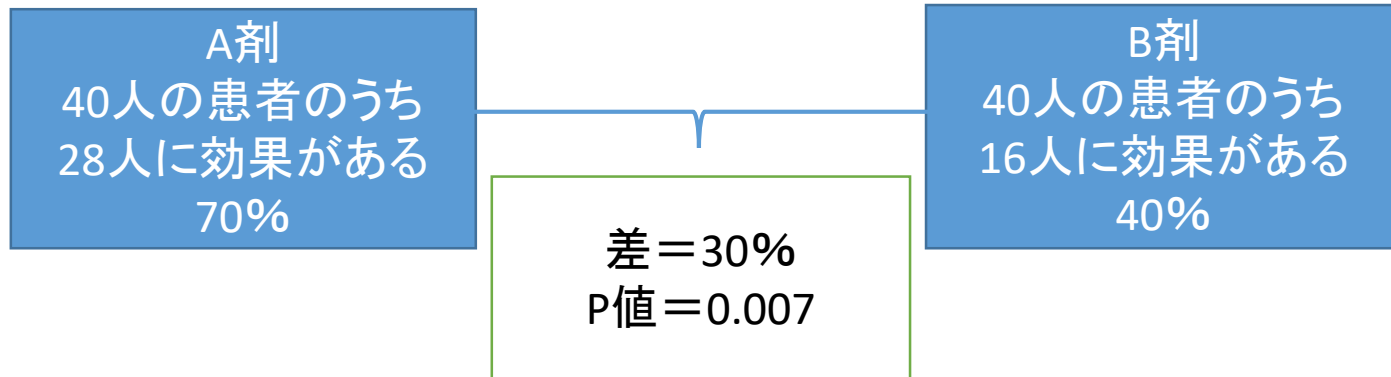
帰無仮説を棄却する場合
その差が有意に認められる

帰無仮説を棄却できない場合
その差が有意に認められない

サンプルサイズとP値



Not significant ($p=0.178$)



Significant ($p=0.007$)

第1種の過誤(α エラー):

差がないのに誤って有意差ありと判定すること→読みすぎ
一般的にはその有意水準(危険率)は5%未満(0.05)と設定する

有意水準が高いほど帰無仮説が棄却されやすくなる

有意水準は一定の場合、サンプルサイズ↑差の有意に検出↑

第2種の過誤(β エラー)

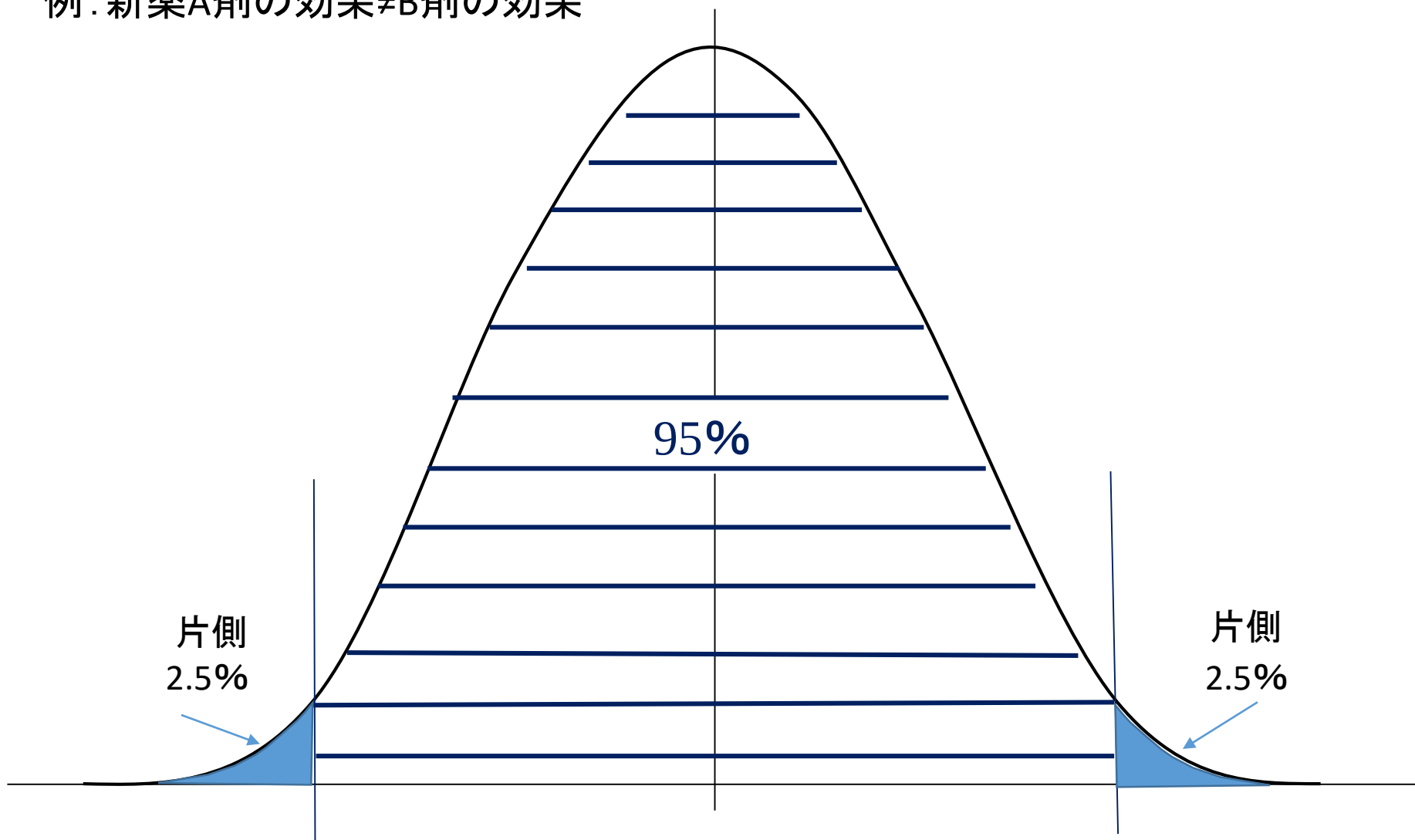
差があるのに誤って差がないと判定すること→見落とし
一般的にはその有意水準(危険率)は、20%(0.2)と設定する
検出力(パワー) = $(1 - \beta) = 80\% (0.8)$ 或いは $90\% (0.9)$
検出力とは、差があるならどの程度で正しく検出することができるか。

有意水準が低いほど帰無仮説が棄却されにくくなる

検出力は一定の場合、サンプルサイズ↑差の有意に検出↑

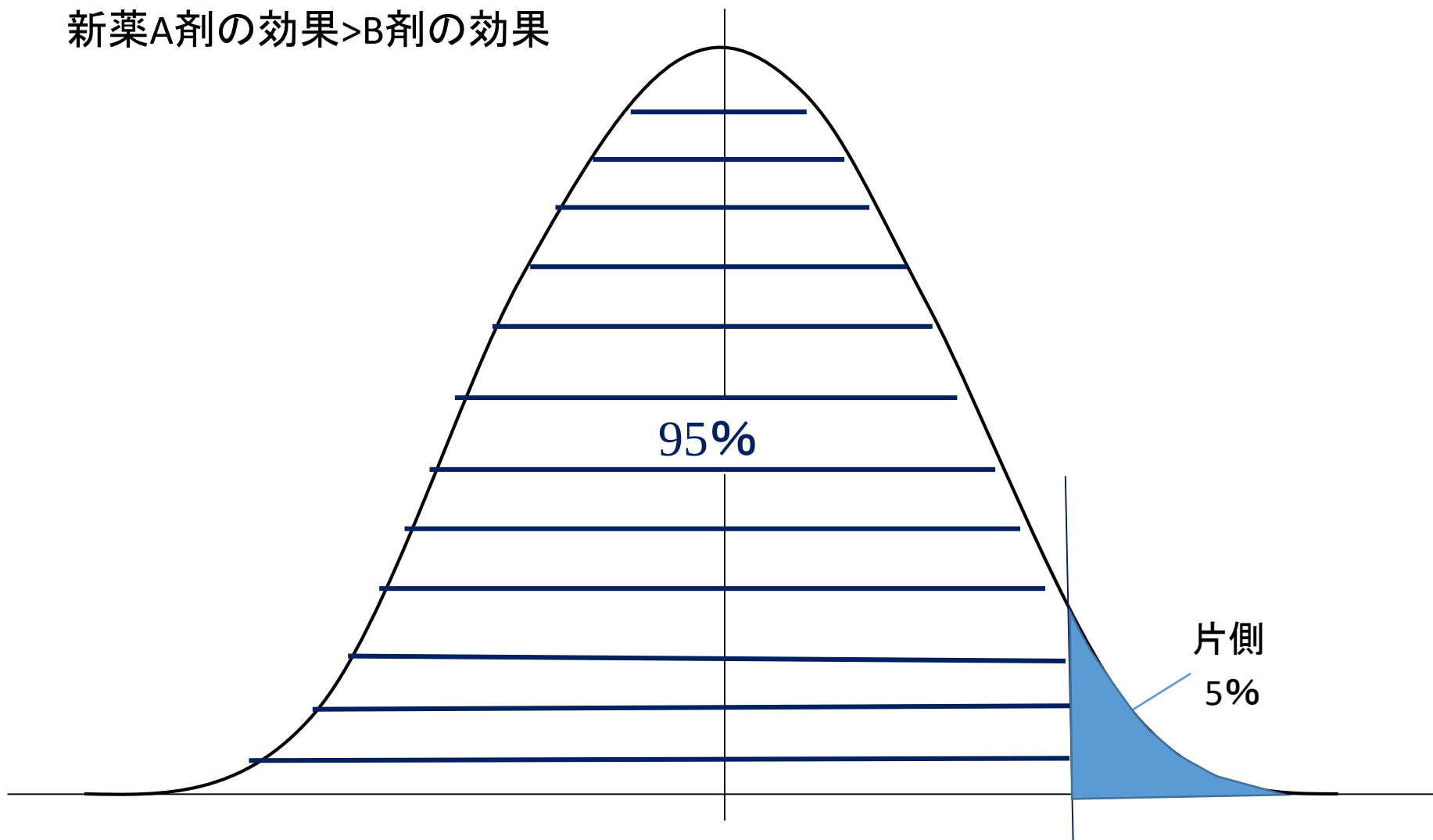
両側検定

例：新薬A剤の効果 $\neq$ B剤の効果



片側検定

例：
新薬A剤の効果>B剤の効果



標本サイズの計算

簡便な公式(有意水準5%,検出力80%,両側検定)

対応のないt検定の場合

臨床的に意味ある平均値の差
(研究期待の差)

観察値の標準偏差
(公表された研究情報の利用)

=標準化された差

$$\text{標本サイズ} = \frac{16}{(\text{標準化された差})^2}$$

カイ二乗検定の場合

臨床的に意味ある2つ群の割合の差
P1-P2

=標準化された差

$$\sqrt{\frac{P1+P2}{2} \left(1 - \frac{P1+P2}{2}\right)}$$

サンプルサイズ計算例(1)

高血圧症患者に対して新しい降圧剤(A剤)を投与し、従来の降圧剤(B剤)と比較したい。平均血圧5mmHgの違いを検証するためには介入群と対照群の必要な人数を求めよ。先行研究の血圧の標準偏差は9mmHgである。(5%の有意水準、80%に検出力、両側検定)

$$\text{標準化された差} = \frac{\begin{array}{c} \text{臨床的に意味ある平均値の差} \\ \text{(研究期待の差)} \end{array}}{\begin{array}{c} \text{観察値の標準偏差} \\ \text{(公表された研究情報の利用)} \end{array}} = \frac{5 \text{ mmHg}}{9 \text{ mmHg}} = 0.556$$

$$\text{標本サイズ} = \frac{16}{(\text{標準化された差})^2} = \frac{16}{0.556^2} = 52 \text{ 人}$$

介入群(A剤)と対照群(B剤)はそれぞれ最低限52人が必要

サンプルサイズ計算例(2)

肩関節周囲炎の治療において、ヒアルロン酸関節内に注射の効果
を物理療法と比較する。臨床的に重要な2群間の治療成功率の差
が25%である。最も治療成績の悪い群の治療成功率を40%とする。
両群の患者数は何人か？(5%有意水準、80%検出力、両側検定)

臨床的に意味ある2つ群の割合の差

$P1-P2$

$0.65-0.40$

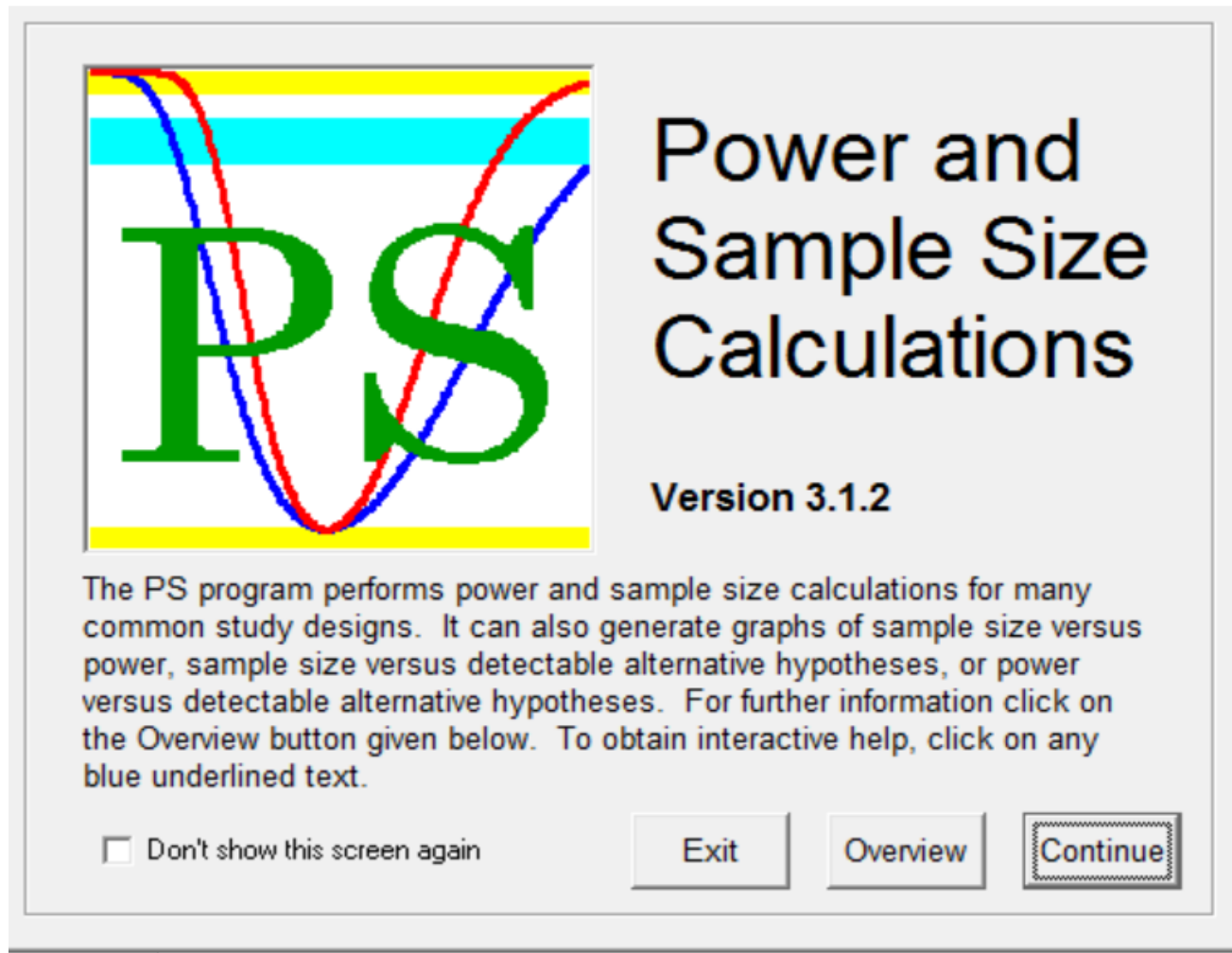
標準化された差 =

$$\frac{P1-P2}{\sqrt{\frac{P1+P2}{2} \left(1 - \frac{P1+P2}{2}\right)}} = \frac{0.65-0.40}{\sqrt{\frac{0.65+0.40}{2} \left(1 - \frac{0.65+0.40}{2}\right)}}$$

$$\text{標本サイズ} = \frac{16}{(\text{標準化された差})^2} = \frac{16}{0.5^2} = 64 \text{ 人}$$

薬物治療群と物理療法群はそれぞれ最低限64人が必要

サンプルサイズ計算ソフトウェア



Department of Biostatistics, Vanderbilt University

<http://biostat.mc.vanderbilt.edu/wiki/Main/PowerSampleSize>

PS Power and Sample Size Program: Main Window

File Edit Log Help

Survival **t-test** Regression 1 Regression 2 Dichotomous Mantel-Haenszel Log

[Studies that are analyzed by t-tests](#)

Output

[What do you want to know?](#)

Sample size
Power
Detectable alternative

Design

[Paired or independent?](#)

Paired
Independent

Input

Input fields will appear when the Output and Design sections are complete

Calculate

Graphs

Description

PS version 3.1.2

Copy to Log Exit

Logging is enabled.

Survival

t-test

Regression 1

Regression 2

Dichotomous

Mantel-Haenszel

Log

Output

[Studies that are analyzed by t-tests](#)[What do you want to know?](#)

Sample size

Design

[Paired or independent?](#)

Independent

Input

 α δ σ m [power](#)

Calculate

Graphs

Description

PS version 3.1.2

Copy to Log

Exit

Logging is enabled.

Input variables on the *t*-test dialog box

α

The Type I error probability for a two sided test. This is the probability that we will falsely reject the null hypothesis.

n

For independent t-tests n is the number of experimental subjects. For pair test n is the number of pairs.

power

For independent tests *power* is probability of correctly rejecting the null hypothesis of equal population means given n experimental patients, m control patients per experimental patient, a Type I error probability α and a true difference in population means of δ . For paired tests it is the probability of correctly rejecting the null hypothesis of equal population means given n pairs of patients, a Type I error probability α and a true difference in population means of δ .

δ

A difference in population means

σ

For independent tests σ is the within group standard deviation. For paired designs it is the standard deviation of difference in the response of matched pairs.

m

For independent tests m is the ratio of control to experimental patients. m is not defined for paired studies.

[Studies that are analyzed by t-tests](#)

Output

[What do you want to know?](#)

Sample size

Design

[Paired or independent?](#)

Independent

Input

 α

0.05

 δ

両群の差

 σ

群のSD

[power](#)

0.8

 m 対照群と実験
群の人数の比

Calculate

Graphs

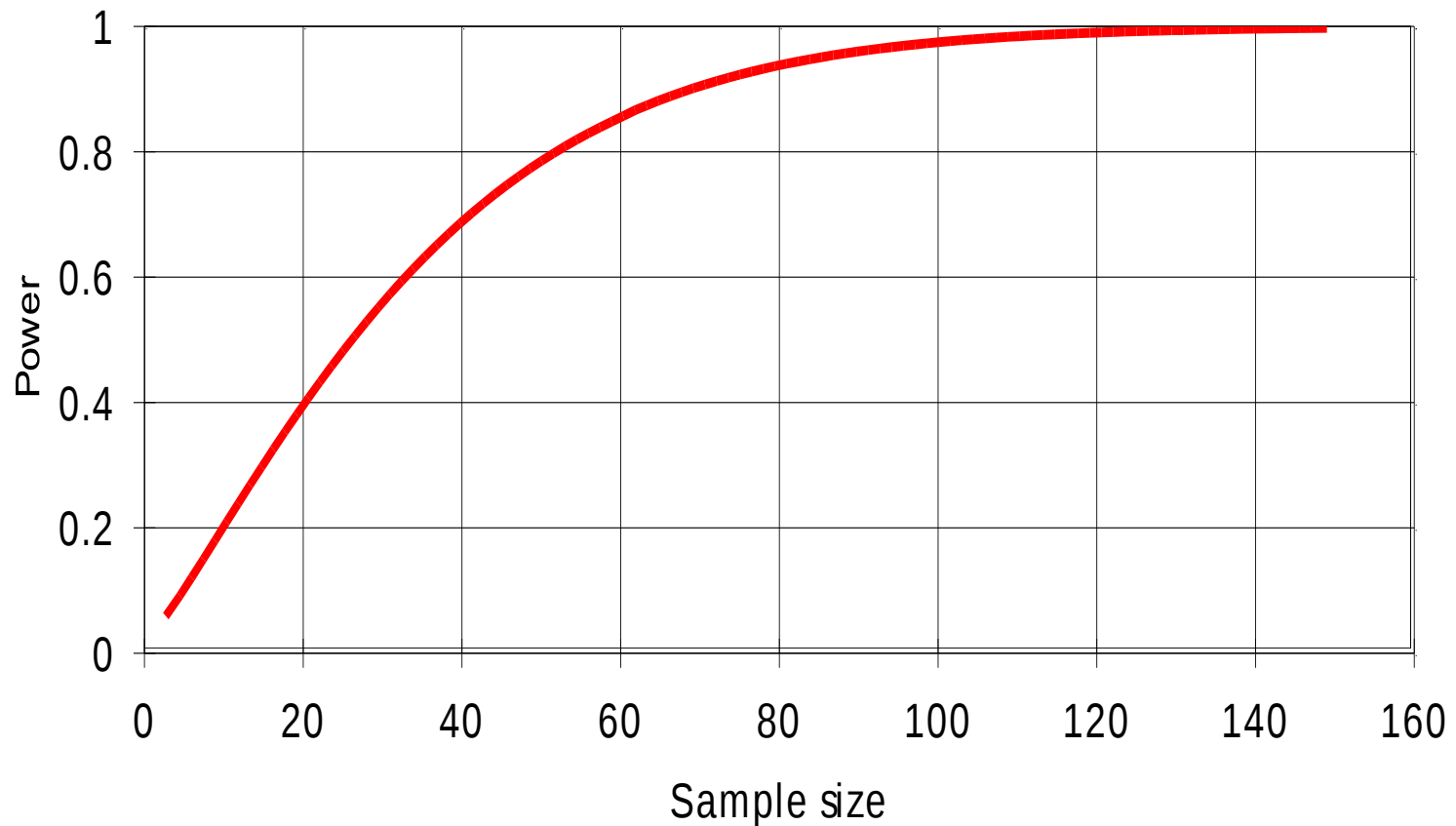
Description

方法の記載(参考)

We are planning a study of a continuous response variable from independent control and experimental subjects with 1 control(s) per experimental subject. In a previous study the response within each subject group was normally distributed with standard deviation 9. If the true difference in the experimental and control means is 5, we will need to study 52 experimental subjects and 52 control subjects to be able to reject the null hypothesis that the population means of the experimental and control groups are equal with probability (power) 0.8. The Type I error probability associated with this test of this null hypothesis is 0.05.

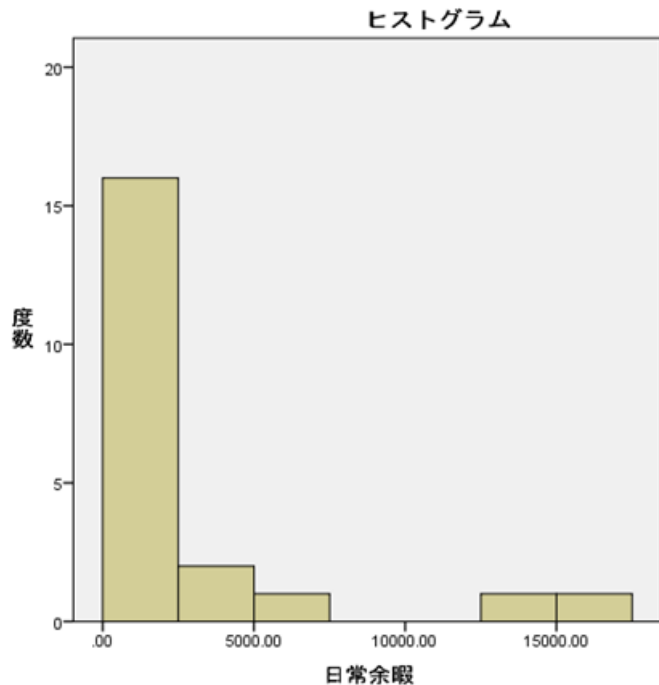
PSソフトウェアの活用

臨床的意味がある差によるサンプルサイズ→パワー

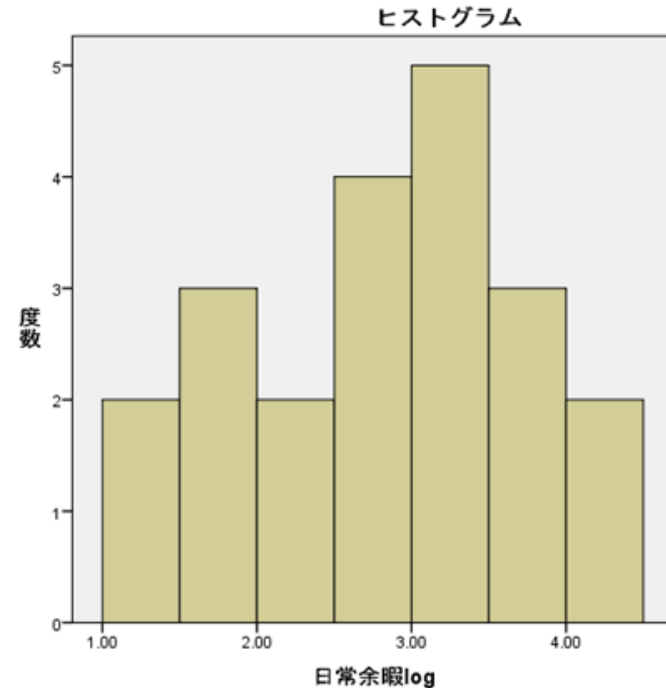


正規分布に近ければ標準偏差(SD)↓サンプルサイズ↓

心筋梗塞患者の日常消費カロリー/週



両群の差 = $3598 - 2598 = 1000$
標準偏差 = 2127
サンプルサイズ = 72



両群の差 = $\log_{10}(3598) - \log_{10}(2598) = 0.14$
標準偏差 = 0.25
サンプルサイズ = 51

[Studies that are analyzed by t-tests](#)

Output

[What do you want to know?](#)

Sample size

Design

[Paired or independent?](#)

Paired

Input

 α

0.05

 δ

前後の差

 σ

前後の差のSD

Calculate

[power](#)

0.8

Graphs

Description

Studies that are analyzed by chi-square or Fisher's exact test

Output

What do you want to know?

Sample size

Case sample size for uncorrected
chi-squared test

61

Design

Matched or Independent?

Independent

Case control?

Case-Control

How is the alternative hypothesis expressed?

Two proportions

Uncorrected chi-square or Fisher's exact test?

Uncorrected chi-square test

Input

 α

0.05

 p_0

0.4

power

0.8

 p_1

0.65

 m

1

Calculate

Graphs

Description

We are planning a study of independent cases and controls with 1 control(s) per case. Prior data indicate that the probability of exposure among controls is 0.4. If the true probability of exposure among cases is 0.65, we will need to study 61 case patients and 61 control patients to be able to reject the null hypothesis that the exposure rates for case and controls are equal with probability (power) 0.8. The Type I error probability associated with this test of this null hypothesis is 0.05. We will use an uncorrected chi-squared statistic to evaluate this null hypothesis.

ケース・コントロール研究

 P_0 : 対照群の曝露因子の割合 P_1 : 実験(観察)群の曝露因子の割合 m : 対照群と実験群の人数比

Prospective(前向き研究)

 P_0 : 対照群のアウトカムの割合 P_1 : 実験(観察)群のアウトカムの割合 m : 対照群と実験群の人数比

Survival | t-test | Regression 1 | Regression 2 | Dichotomous | Mantel-Haenszel | Log

Output

[Studies that are analyzed by log-rank tests](#)[What do you want to know?](#)

Sample size

[Sample Size](#)

118

Design

[How is the alternative hypothesis expressed?](#)

two survival times

Hazard ratio or relative risk

Input

 α

0.05

 A

24

Calculate

 $power$

0.8

 m_1

24

 F

60

Graphs

 m_2

36

 m

1

Description

We are planning a study with 1 control per experimental subject, an accrual interval of 24 time units, and additional follow-up after the accrual interval of 60 time units. Prior data indicate that the median survival time on the control treatment is 24 time units. If the true median survival times on the control and experimental treatments are 24 and 36 time units, respectively, we will need to study 118 experimental subjects and 118 control subjects to be able to reject the null hypothesis that the experimental and control survival curves are equal with probability (power) .800. The Type I error probability

m_1 : 対照群の生存
期間の中央値

m_2 : 実験(観察)群の生存
期間の中央値

A : 患者をリクルートする期間

F : リクルートした最後の患者
の後の追跡期間

m : 対照群と実験群の人数比

R : 実験群と対照群のハザード
比(相対リスク)。ハザード比
(相対リスク)は一定の場合、
 $R = m_2/m_1$

注意事項
各時間の単位を統一する必要

統計ソフトについて EZRのインストール



自治医科大学附属さいたま医療センター
Saitama Medical Center Jichi Medical University

[ホーム](#)[施設紹介](#)[移植診療](#)[受診の案内](#)[研修・研究](#)[統計ソフトEZR](#)[フリー統計ソフトEZR](#)[EZRの解析機能](#)[EZRの使い方、変更履歴](#)[ダウンロード \(Windows標準版\)](#)[ダウンロード \(MacOS X版\)](#)[ダウンロード \(LINUX版\)](#)[よくあるご質問 \(FAQ\)](#)[English version](#)

***** [Click here for English version](#) *****

256 136

(2012年6月11日のページ改訂後のアクセス数です。)

2016年2月3日 EZR version 1.32公開

2014年11月初心者向けマニュアル刊行

2015年4月EZRマニュアル第2版刊行

1.23からサンプルの背景データのサマリー表(Table 1)を自動作成する新機能を搭載!!

動作確認済OS Windows XP~10、Mac OS X Snow Leopard~El Captain、Ubuntu 11.10~15.10

EZRを使用した学术论文を発表される場合は

[Bone Marrow Transplantation 2013; 48, 452-458](#)

を参考文献として引用くださいますようお願いいたします。

(「EZRの使い方、変更履歴」を参照。既に200編以上の英文論文に引用されています)

国際的に有名な[Softpedia software database](#)に

"100% CLEAN" Softpedia awardで掲載されました。

フリー統計ソフトEZR(Easy R)とは？

EZRと統計の解説書



初心者でもすぐにできる
フリー統計...
神田 善伸

ロープライス ¥4,104
or 新品 ¥4,104
ポイント 123pt

[amazon.co.jp](#)
で買う

プライバシーについて

↑ 初級～中級者用



EZRでやさしく学ぶ統計
学～EBMの真実...
神田 善伸

ロープライス ¥4,968
or 新品 ¥4,968
ポイント 150pt

[amazon.co.jp](#)
で買う

プライバシーについて

<http://www.jichi.ac.jp/saitama-sct/SaitamaHP.files/statmed.html>

データセット

クリプト Rマ

- 名義変数の解析 ▶
- 連続変数の解析 ▶
- ノンパラメトリック検定 ▶
- 生存期間の解析 ▶
- 検査の正確度の評価 ▶
- マッチドペア解析 ▶
- メタアナリシスとメタ回帰 ▶
- 必要サンプルサイズの計算 ▶

編集

表示

保存

モデル :

Σ <アクティブモデルなし

閾値奏効率、期待奏効率からのサンプルサイズの計算

1群の比率の信頼区間をある幅におさめるためのサンプルサイズの計算

1群の比率を既知の比率と比較するためのサンプルサイズの計算

1群の比率を既知の比率と比較するための検出力の計算

2群の比率の比較のためのサンプルサイズの計算

2群の比率の比較のための検出力の計算

2群の比率の比較(非劣性)のためのサンプルサイズの計算

1群の平均値の信頼区間をある幅におさめるためのサンプルサイズの計算

2群の平均値の比較のためのサンプルサイズの計算

2群の平均値の比較のための検出力の計算

2群の平均の比較(非劣性)のためのサンプルサイズの計算

対応のある2群の平均値の比較のためのサンプルサイズの計算

対応のある2群の平均値の比較のための検出力の計算

2群の生存曲線の比較のためのサンプルサイズの計算

2群の生存曲線の比較のための検出力の計算

2群の生存曲線の比較(非劣性)のためのサンプルサイズの計算

サンプルサイズに関するいくつかの注意点

- 研究倫理の遵守
- サンプルサイズの計算は臨床研究の前に行う
- 計算したサンプルサイズ＋脱落者(～20%)で見積り
- 計算ソフトウェアによってサンプルサイズは少し違う場合がある
- 臨床研究の終了の前には差があるかどうかを解析しないのは一般的
- サンプルサイズは少ない(多い)が、有意な結果が出た場合の対応
- 特殊のサンプルサイズの計算は気軽にご相談ください。

End

