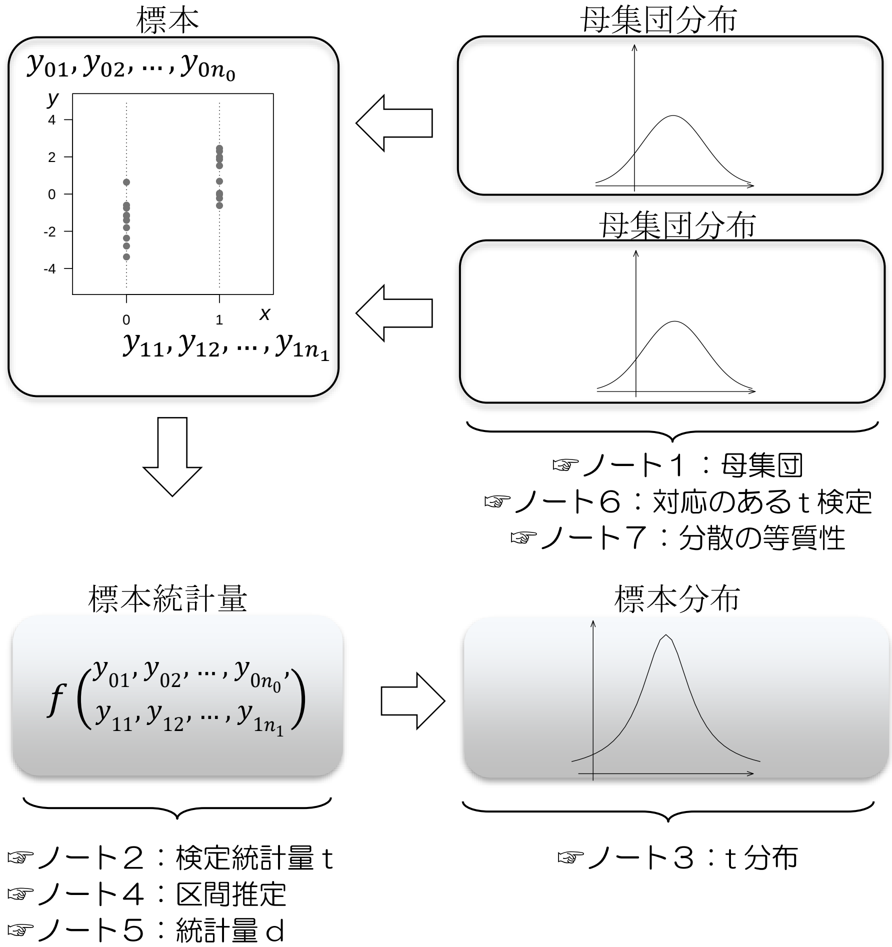
第3講　二群の差の検定（対応のない t 検定）

学習の目標

- □ 二群の平均値の差の検定である「対応のない t 検定」のロジックがわかり、論文で出会った際に適切に解釈できる。
- □ t 検定を行う際に母集団に想定される仮定を理解できる。
- □ 仮説検定における二つの仮説（帰無仮説と対立仮説）が理解でき、適切に想定できる。
- □ 標本分布の標準偏差を標本誤差と呼び、標本平均値差が従う標本分布の標準誤差を$\sigma_{\bar{y}_1-\bar{y}_0}$と書くことが分かる。
- □ t 値という統計量は、標本平均値差とその標本誤差の推定量の比の値であることが分かる。
- □ 帰無仮説の下で t 値が従う標本分布を t 分布と言い、自由度が大きくなると標本正規分布に近づくことが分かる。
- □ t 検定では、データから得られた t 値が、限界値を超えて t 分布の棄却域に入るかどうかを検定する。棄却域に入った場合、帰無仮説という仮定が誤っていたと棄却する。
- □ t 値が棄却域に落ちる確率を有意水準 α と呼び、帰無仮説が真の時、誤って棄却する第一種の過誤を表すことが分かる。
- □ 対立仮説が正しい時に t 値が従う分布を非心 t 分布と呼び、これに基づいて対立仮説が正しいのに帰無仮説を保持してしまう誤りを β（第二種の過誤）と呼ぶことが分かる。
- □ 信頼区間の作り方がわかり、95%信頼区間とは「標本から計算された区間が母集団のパラメータ値を含む確率」が 95%であることを意味するということが分かる。
- □ サンプルサイズをあまりに大きくすると、細かい差まで際限なく検出できるようになり、この結果実質的には小さな差しかないのに有意になってしまうことが分かる。
- □ 平均値の差が標準偏差が何個分なのかを表す統計量としてコーエンの d という効果量があることが分かる。
- □ 二種類の過誤と母集団効果量からサンプルサイズの決定に利用できる。

本講から数回にわたり、t 検定と呼ばれる統計手法を学びます。これから学んでいく統計手法の中ではもっとも単純なモデルの一つで、論文などでも目にすることの多い伝統的な統計手法です。しかし「単純」なモデルではありますが、より発展的な手法の基礎となるモデルでもあります。しっかりと身に付けてください。

見取り図



データの形式

ID	グループ（説明変数）	観測値（応答変数）
1	0	2.1
2	0	3.2
⋮	⋮	⋮
n_0	0	3.1
n_0+1	1	2.3
n_0+2	1	1.2
⋮	⋮	⋮
n_0+n_1	1	1.5

（1）目的（リサーチクエスチョン）

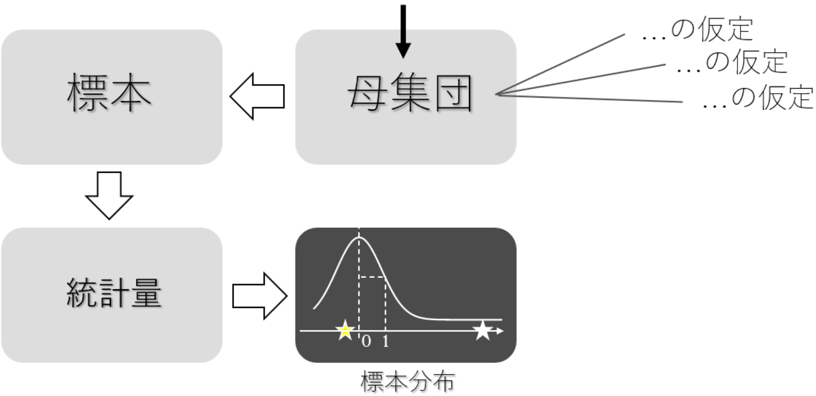
標本が採られてきていると想定される二種類の母集団の母平均（期待値）に差があるか否か、判断を下す二群の差の検定を学ぶ。

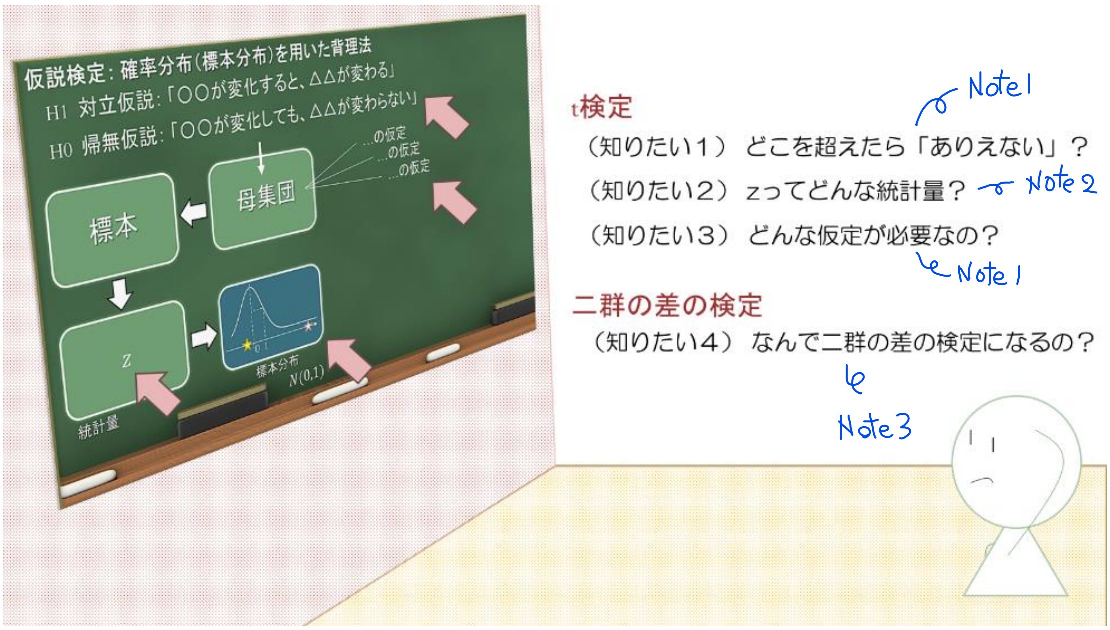
（2）考え方

仮説検定： 確率分布（標本分布）を用いた背理法

H1 対立仮説： 「〇〇が変化すると、△△が変わる」

H0 帰無仮説： 「〇〇が変化しても、△△が変わらない」





📖 ノート１　仮説検定

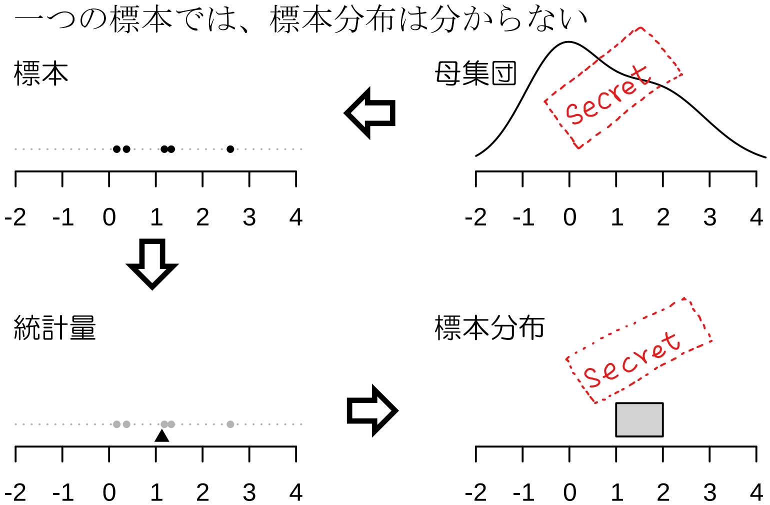
○ 仮説検定

一回しか標本を採っていない（真の標本分布は分からない）のに、①条件さえ整えば標本分布の姿を特定できる統計量があることに注目し、②母集団に対する帰無仮説を、③その統計量の標本分布を用いた背理法で棄却しようとする試み。

（１）ポイント１：標本分布を特定する

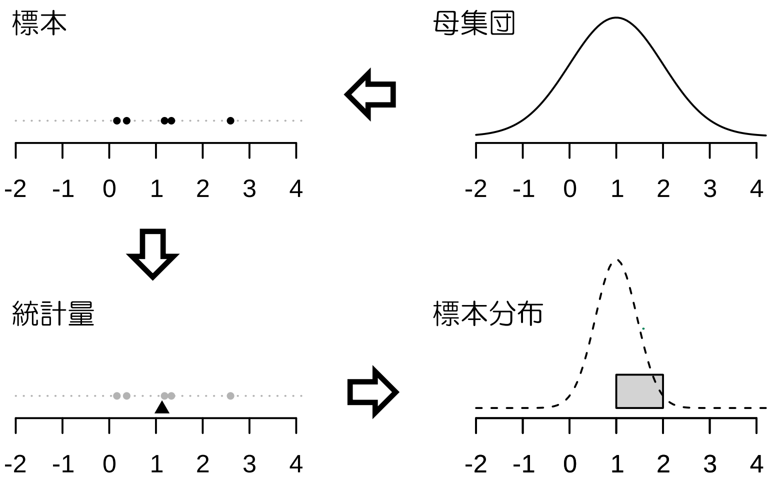
① 普通の統計量

一つの標本では、標本分布は分からない



② 特殊な統計量　→ 中心極限定理（第2講）

数学的な性質が解明されている一部の統計量は、条件さえ整えば、標本を採らずとも、標本分布が分かる



Ⓟ 独立性の仮定が成り立たない時

ID		英語
1	山田	30
2	山田	40
3	田中	90
4	田中	80
⋮		⋮

同じ人物が反復して観測されたことで、
ID＝1のデータと
ID＝2のデータに
関連性が生じてしまった
(＝独立ではない)
→ 第7講で扱う手法で対応

（２）ポイント２：母集団への仮説

① 帰無仮説 Null Hypothesis (H0)
棄却する目的で設定される仮説 ←「ニ」

② 対立仮説 Alternative Hypothesis (H1)
帰無仮説が棄却された際に採択する仮説 ←「キ」

その他に必要な標本分布を同定するための仮定

【標本の抽出の仕方に関する仮定】

independently	① 独立性の仮定：	各要素は互いに独立
identical	② 同一分布の仮定：	標本は同じ分布から抽出
	③ 無作為性の仮定：	標本はランダムに抽出

$$y_1, y_2, \cdots, y_n \overset{i.i.d}{\sim} N(\mu, \sigma^2)$$

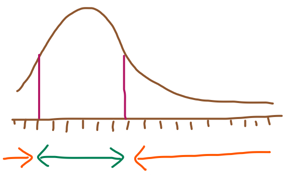
【母集団の姿に関する仮定】

例：t 検定の場合

① 正規性の仮定：	母集団の分布は正規分布
② 分散の等質性の仮定：	二つの母集団の分散は同じ

Ⓟ 棄却/保持すること

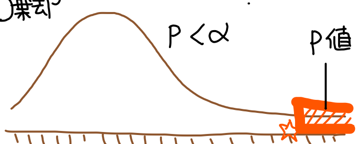
(1) 領域で考える(数直線で表現)



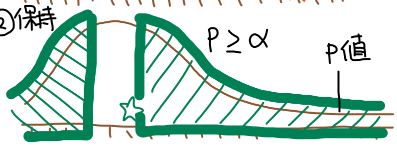
☆が受容域に入ったのか
棄却域に入ったのかで表現

(2) 面積で考える(確率で表現)

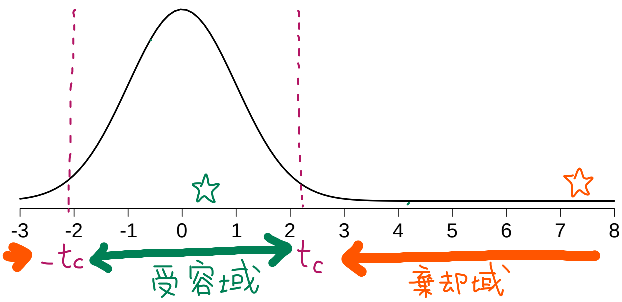
①棄却



②保持



（３）ポイント３：標本分布を用いた背理法



① 限界値（臨界値）t_c
よくあることと滅多にないことを分ける基準

② p 値
帰無仮説が正しい時に、今回の検定統計量の実現値が、今以上に極端になる確率のこと

③ 棄却域と受容域
棄却域：限界値より極端な値を示した領域
受容域：棄却域ではない領域

④ 有意水準α
統計量が棄却域に入る確率のこと

📖 ノート２　標本分布１：標準正規分布から t 分布へ

Ⓟ 数式に強くなるコツ

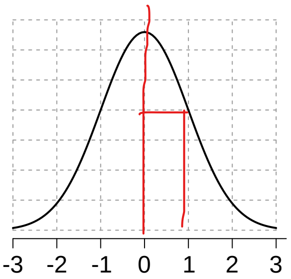
$$z_i = \frac{\boxed{データ} - \boxed{中心}}{\boxed{ばらつきの単位}}$$

（ステップ1）引き算
中心からデータがどれくらい離れているのか

（ステップ2）割り算
それがばらつきの単位何個分か。

（１）標準正規分布 Standard Normal Distribution

これは、平均が 0、分散が 1 の正規分布$N(0,1)$。

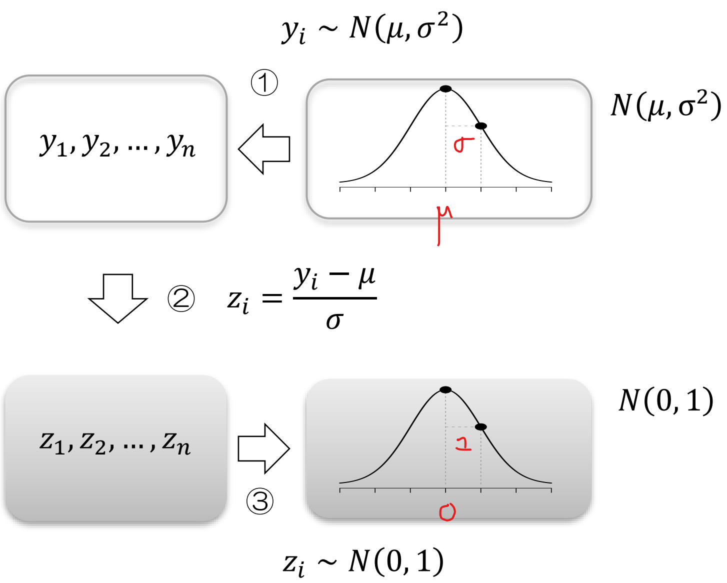


（２）標準化 Standardization

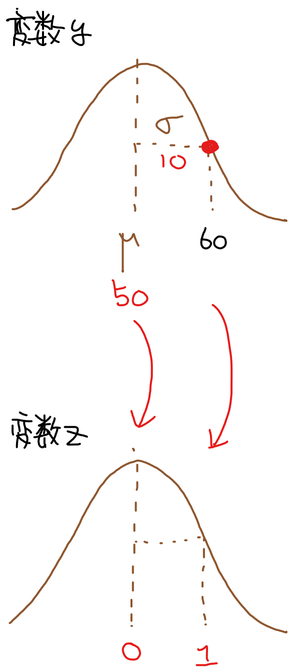
これは、「分布の中心」から「標準偏差何個分なのか」を表すように、変数を変換すること。

$$z_i = \frac{y_i - \mu}{\sigma}$$

変数y_iが正規分布$N(\mu, \sigma^2)$に従うとき、標準化という操作で新しく作られた新しい統計量z_iは標準正規分布に従う。

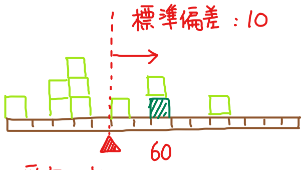


Ⓟ 理由



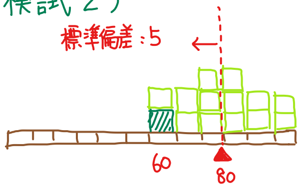
Ⓟ 偏差値

(模試1)



A君は, 平均(60)から標準偏差(10)が +1個の位置にいる

(模試2)



B君は, 平均(80)から標準偏差(5)が −4個の位置にいる

「標準化得点」

偏差値

① 例題

関東の予備校の英語の模試で 60 点を取った A 君と関西の予備校の英語の模試で 60 点を取った B 君はどちらが英語の成績が良いと言えるのだろうか？

② 偏差値

異なる標本間での点数の比較のために各データが「平均から標準偏差何個分」の位置にあるか計算したもの。

$$z_i = \underbrace{\left[50 + 10 \times \underbrace{\left(\frac{\overset{\text{60点}}{y_i} - \overset{\text{50点}}{\bar{y}}}{\underset{\text{10点}}{s_y}} \right)}_{\text{標準化得点}} \right]}_{\text{偏差値}}$$

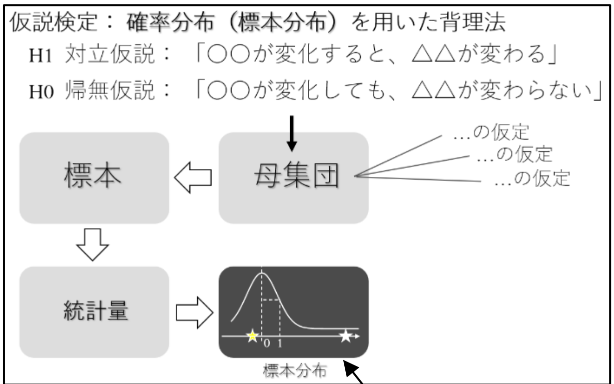
※ただし、もとのテストの点数と似たような値に収まるよう 10 をかけて 50 を足している。

③ 標準化

異なる標本間での点数の比較のために各データが「平均から標準偏差何個分」の位置にあるか計算したもの。

$$z_i = \frac{y_i - \bar{y}}{s_y}$$

ねらい

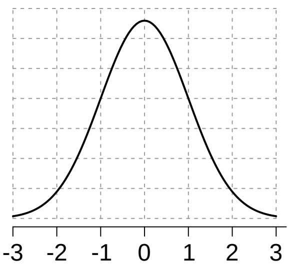


できることなら、ここに**標準正規分布**を使いたい。

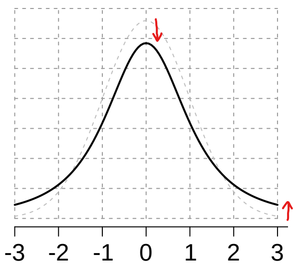
（3）t 分布　t Distribution

これは、σではなく、$\hat{\sigma}$を用いるという補正された標準化得点（統計量）がしたがう、標準正規分布にな̀り̀か̀け̀た̀分布。

標準正規分布 $N(0,1)$



t 分布（例：$t(2)$）



※ 自由度：どれくらい標準正規分布に近いかを表すパラメータ。値が大きいほど、$N(0,1)$に近くなる。

（4）t 値　t value

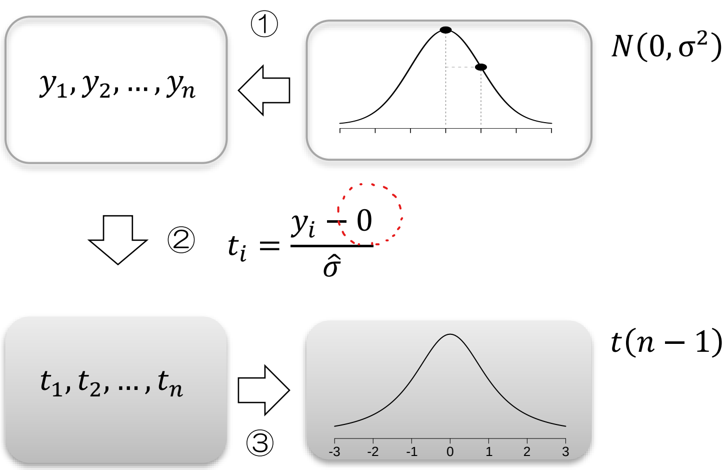
これは、「分布の中心」から「標準偏差何個分なのか」を表す標準化得点の計算過程で、σではなく、$\hat{\sigma}$を用いたもの。

$$t_i = \frac{y_i - \mu}{\hat{\sigma}}$$

μについては、帰無仮説で μ=0を想定する

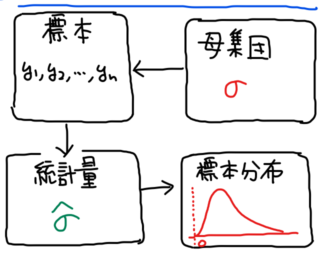
変数y_iが正規分布$N(\mu, \sigma^2)$に従うとき、補正された標準化という操作で作られた統計量t_iは $t(n-1)$ に従う。

$$y_i \sim N(0, \sigma^2)$$



$$t_i \sim t(n-1)$$

Ⓟ 記法：＾（ハット）



$\hat{\sigma}$：母集団におけるσという値を人間が推定した統計量

σ：母集団における定̀数̀

（＝どんな標本をとるのかにかかわらずいつも一定）

Ⓟ t値とz値の差

μ=0 という帰無仮説のもとでは，…

(1) z 値

$$z_i = \frac{y_i}{\sigma}$$

(2) t 値

$$t_i = \frac{y_i}{\hat{\sigma}}$$

となる．

⇒ したがって，$\hat{\sigma}$の精度が良くなればなるほど

t_iはz_iに近づいてゆく

⇒ $\hat{\sigma}$はサンプルサイズが大きいと精度を増す

⇒ サンプルサイズが大きいとt分布は標準正規分布に近づいていく．

Ⓟ推定量 Estimator

① 統計量

これは,標本にもとづいて計算する量のこと

→無数に作れる

(例) $5\times x_1 + x_2 + 19\times x_3$

↳ 5月19日統計量

⇒統計量ではあるが,用途がない

② 推定量

これは,母集団の値の推定に使えるという用途をもつ統計量

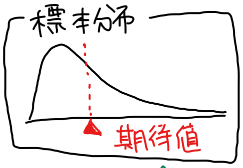
※推定量の良さを測かる基準を考える

(全知全能の視点)

母集団の値に一致するという基準で選べばいい

(人間の視点)

母集団の値に平均的に一致するという基準を使う

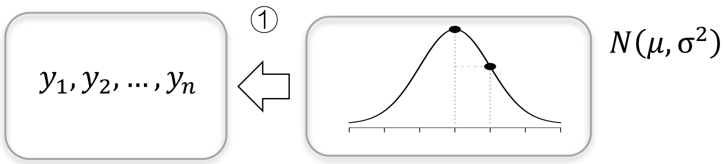


不偏性 : $\sigma = E[\hat{\sigma}]$

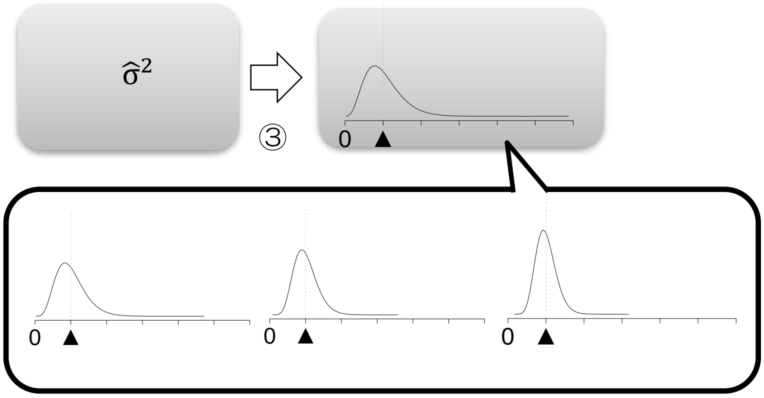
他にもいろいろな基準がある

推定された分散の精度

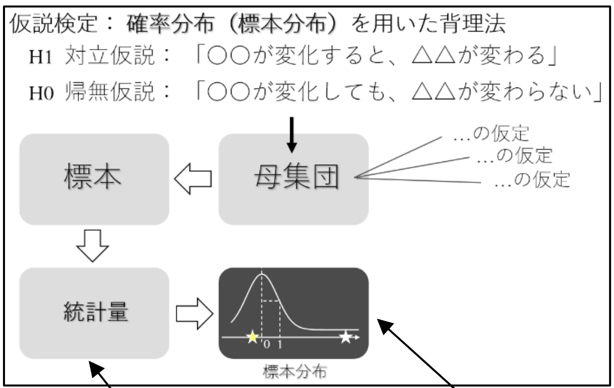
サンプルサイズが大きくなれば、大きくなるほど、分散の推定精度は上がる。



② $$\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$$: 不偏分散



ねらい



$\mu \neq 0$

$\mu = 0$

人間が計算ができるようにするため、z から t に変更。

統計量が変更されたので標準正規分布から t 分布に変化。

（１）母集団の平均が０であることの検定

定数であれば、0でなく2もいい

■ 帰無仮説 H0

標本がしたがう母集団（正規分布）の中心　μ は 0。

Step 1　正規分布にしたがう統計量を見つける

$$y_i \Leftarrow N(\mu, \sigma^2)$$

Step 2　標準化（z 値を考える）

$$y_i \Leftarrow N(\mu, \sigma^2)$$

↓標準化

$$\frac{y_i - \mu}{\sigma} \Rightarrow N(0, 1)$$

Step 3　帰無仮説を導入

$$\frac{y_i - \mu}{\sigma} \Leftarrow N(0, 1)$$

↓帰無仮説（$\mu = 0$）

$$\frac{y_i}{\sigma} \Rightarrow N(0, 1)$$

Step 4　補正された標準化（t 値を考える）

$$\frac{y_i}{\sigma} \Leftarrow N(0, 1)$$

↓点推定

$$t_i = \frac{y_i}{\hat{\sigma}} \Rightarrow t(n-1)$$

Ⓟリサーチクエスチョンの例

A組のクラスの男子30人を標本に採った。この男子たちの身長を計測し，母集団における全ての男子たちの身長の平均が170cmなのか調べたい

H0：$\mu = 170$ cm

H1：$\mu \neq 170$ cm

自由度

今回の t 値である下記の量が$t(n-1)$に従うことの証明

$$t_i = \frac{y_i - \mu}{\hat{\sigma}}$$

【証明】次の F 値が自由度$1, n-1$の F 分布に従うことを示す

$$F_i = t_i^2 = \left(\frac{y_i-\mu}{\hat{\sigma}}\right)^2$$

$$= \left(\frac{y_i-\mu}{\sigma}\right)^2 \bigg/ \left(\frac{\hat{\sigma}}{\sigma}\right)^2$$

☜ 分子と分母に$\frac{1}{\sigma^2}$をかける

$$(\text{分子}) = \left(\frac{y_i-\mu}{\sigma}\right)^2 \sim \chi^2(1)$$

$$= \{\text{自由度 1 の}\chi^2\text{分布に従う変数}\}/1$$

$$(\text{分母}) = \frac{\hat{\sigma}^2}{\sigma^2}$$

$$= \frac{\frac{1}{n-1}\sum_{i=1}^{n}(y_i-\bar{y})^2}{\sigma^2}$$

$$= \frac{\sum_{i=1}^{n}(y_i-\bar{y})^2}{\sigma^2}\bigg/(n-1)$$

$$= \frac{\sum_{i=1}^{n}\{(y_i-\mu)-(\bar{y}-\mu)\}^2}{\sigma^2}\bigg/(n-1)$$

$$= \frac{\sum_{i=1}^{n}\{(y_i-\mu)^2-2(y_i-\mu)(\bar{y}-\mu)+(\bar{y}-\mu)^2\}}{\sigma^2}\bigg/(n-1)$$

$$= \frac{\sum_{i=1}^{n}(y_i-\mu)^2-2\sum_{i=1}^{n}(y_i-\mu)(\bar{y}-\mu)+\sum_{i=1}^{n}(\bar{y}-\mu)^2}{\sigma^2}\bigg/(n-1)$$

$$= \left\{\frac{\sum_{i=1}^{n}(y_i-\mu)^2}{\sigma^2} - \frac{2(\bar{y}-\mu)\sum_{i=1}^{n}(y_i-\mu)}{\sigma^2} + \frac{n(\bar{y}-\mu)^2}{\sigma^2}\right\}\bigg/(n-1)$$

$$= \left\{\sum_{i=1}^{n}\frac{(y_i-\mu)^2}{\sigma^2} - \frac{2(\bar{y}-\mu)n(\bar{y}-\mu)}{\sigma^2} + \frac{n(\bar{y}-\mu)^2}{\sigma^2}\right\}\bigg/(n-1)$$

$$= \left\{\sum_{i=1}^{n}\left(\frac{y_i-\mu}{\sigma}\right)^2 - \frac{2n(\bar{y}-\mu)^2}{\sigma^2} + \frac{n(\bar{y}-\mu)^2}{\sigma^2}\right\}\bigg/(n-1)$$

$$= \left\{\sum_{i=1}^{n}\left(\frac{y_i-\mu}{\sigma}\right)^2 - \frac{n(\bar{y}-\mu)^2}{\sigma^2}\right\}\bigg/(n-1)$$

$$= \left\{\sum_{i=1}^{n}\left(\frac{y_i-\mu}{\sigma}\right)^2 - \left(\frac{\bar{y}-\mu}{\sigma/\sqrt{n}}\right)^2\right\}\bigg/(n-1)$$

標準化得点$\sim \chi^2(n)$ ↑　　　　↑ 標準化得点 $\sim \chi^2(1)$

$$= \{\text{自由度 n}-1\text{ の}\chi^2\text{分布に従う変数}\}/(n-1)$$

$$\therefore\ F_i = t_i^2 = \frac{\{\text{自由度 1 の}\chi^2\text{分布に従う変数}\}/1}{\{\text{自由度}n-1\text{ の}\chi^2\text{分布に従う変数}\}/(n-1)} \sim F(1, n-1)$$

0以外の定数でもよい

（2）平均値が０であることの検定

■ 帰無仮説 H0

標本平均がしたがう標本分布（正規分布）の中心 μ は 0。

Step 1 正規分布にしたがう統計量を見つける

$y_i \Leftarrow N(\mu, \sigma^2)$

↓中心極限定理

$\bar{y} \Leftarrow N\left(\mu, \frac{\sigma^2}{n}\right)$

Step 2 標準化（z 値を考える）

$\bar{y} \Leftarrow N\left(\mu, \frac{\sigma^2}{n}\right)$

↓ 標準化

$\frac{\bar{y}-\mu}{\sqrt{\frac{\sigma^2}{n}}} \Leftarrow N(0,1)$

Step 3 帰無仮説を導入

$\frac{\bar{y}-\mu}{\sqrt{\frac{\sigma^2}{n}}} \Leftarrow N(0,1)$

↓帰無仮説（$\mu = 0$）

$\frac{\bar{y}}{\sqrt{\frac{\sigma^2}{n}}} \Leftarrow N(0,1)$

Step 4 補正された標準化（t 値を考える）

$\frac{\bar{y}}{\sqrt{\frac{\sigma^2}{n}}} \Leftarrow N(0,1)$

↓ 点推定

$\frac{\bar{y}}{\sqrt{\frac{\hat{\sigma}^2}{n}}} \Leftarrow t(n-1)$

Ⓟ リサーチクエスチョンの例

男子30人のA組の身長のクラス平均を出してみた。

でも、クラス平均は、クラスや学校によってばらばらだろう。

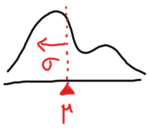
そこで、世の中の全てのクラスの平均値の平均が 170cm なのかを検証する。

H0 : μ = 170 cm

H1 : $\mu \neq$ 170 cm

Ⓟ 中心極限定理

(バージョン1) 任意の母集団



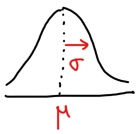
平均が μ で，分散が σ^2 の任意の母集団分布から

$y_1, y_2, \cdots, y_n$ を取りだし，標本平均 $\bar{y}$ を計算すると，

「サンプルサイズが大きいとき，」

$\bar{y} \sim N(\mu, \frac{\sigma^2}{n})$

(バージョン2) 正規分布の母集団



平均が μ で，分散が σ^2 の正規分布の母集団分布から

$y_1, y_2, \cdots, y_n$ を取りだし，標本平均 $\bar{y}$ を計算すると，

「サンプルサイズが大きくなくても」

$\bar{y} \sim N(\mu, \frac{\sigma^2}{n})$

自由度

今回の t 値である下記の量が$t(n-1)$に従うことの証明

$$t = \frac{\bar{y} - \mu}{\hat{\sigma}/\sqrt{n}}$$

【証明】次の F 値が自由度$1, n-1$の F 分布に従うことを示す

$$F = t^2 = \left(\frac{\bar{y}-\mu}{\hat{\sigma}/\sqrt{n}}\right)^2$$

$$= \left(\frac{\bar{y}-\mu}{\sigma/\sqrt{n}}\right)^2 \Bigg/ \left(\frac{\hat{\sigma}}{\sigma}\right)^2$$

$$(\text{分子}) = \left(\frac{\bar{y}-\mu}{\sigma/\sqrt{n}}\right)^2 \sim \chi^2(1)$$

$$= \{\text{自由度 } 1 \text{ の} \chi^2 \text{分布に従う変数}\}/1$$

$$(\text{分母}) = \frac{\hat{\sigma}^2}{\sigma^2}$$

※ 一つ前と同じ計算になる

$$= \left\{\sum_{i=1}^{n}\left(\frac{y_i-\mu}{\sigma}\right)^2 - \left(\frac{\bar{y}-\mu}{\sigma/\sqrt{n}}\right)^2\right\}/(n-1)$$

$$= \{\text{自由度} n-1 \text{ の} \chi^2 \text{分布に従う変数}\}/(n-1)$$

$$\therefore\ F_i = t_i^2 = \frac{\{\text{自由度 } 1 \text{ の} \chi^2 \text{分布に従う変数}\}/1}{\{\text{自由度} n-1 \text{ の} \chi^2 \text{分布に従う変数}\}/(n-1)} \sim F(1, n-1)$$

分布同士の関係

正規分布 → 標準正規分布

t分布 → 標準正規分布

標準正規分布 → χ^2分布

χ^2分布 → F分布

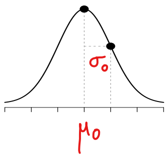
t分布 ⇄ F分布

（３）二群の平均値差の検定

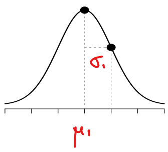
> ■ 帰無仮説 HO
>
> 二つのグループの標本平均の差がしたがう標本分布（正規分布）の中心 $\mu_1 - \mu_0$ は 0。

Step 1 正規分布にしたがう統計量を見つける

$y_{01}, y_{02}, \cdots, y_{0n_0}$ ⇦ $N(\mu_0, \sigma_0^2)$



$y_{11}, y_{12}, \cdots, y_{1n_1}$ ⇦ $N(\mu_1, \sigma_1^2)$



↓中心極限定理

$\bar{y}_0$ ⇦ $N\left(\mu_0, \frac{\sigma_0^2}{n_0}\right)$

＋

$\bar{y}_1$ ⇦ $N\left(\mu_1, \frac{\sigma_1^2}{n_1}\right)$

↓再生性

$\bar{y}_1 - \bar{y}_0$ ⇦ $N\left(\mu_1 - \mu_0, \frac{\sigma_0^2}{n_0} + \frac{\sigma_1^2}{n_1}\right)$

Ⓟ リサーチクエスチョンの例

（事例１）男女による身長の差

① グループ1：男の子

μ_1：男の子の母集団の身長の中心

② グループ0：女の子

μ_0：女の子の母集団の身長の中心

$\mu_1 - \mu_0$（男の子の母集団の身長の平均から女の子の母集団の身長の平均を引いた差）が0なのか？

H0：$\mu_1 - \mu_0 = 0$

H1：$\mu_1 - \mu_0 \neq 0$

（事例２）言語表現の容認度の差

① グループ1：日本語母語

μ_1：日本語母語の人の回答の母集団の平均

② グループ0：非日本語母語

μ_0：非日本語母語の人の回答の母集団の平均

Ⓟ 記法：y_{0i}

二つのグループを話題にしているので、どちらのグループの標本なのかを明示的に表わす

y_{0i}

i 番目だよ

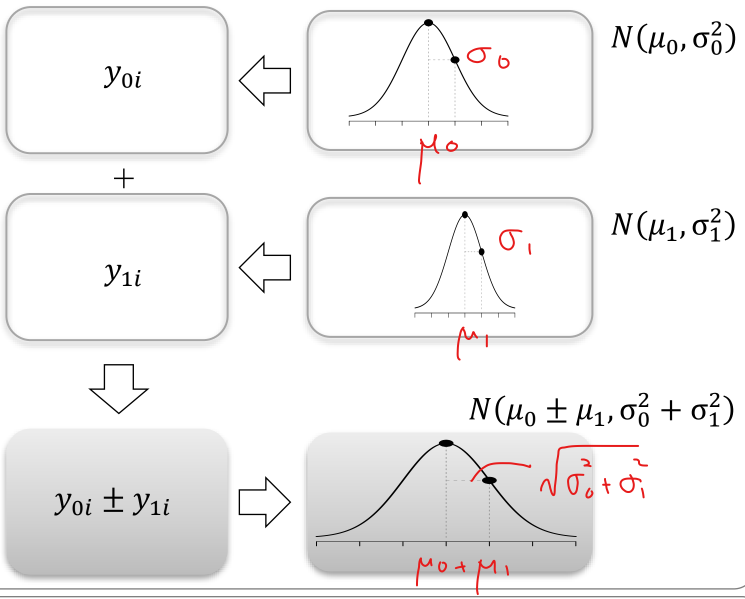
グループ0だよ

再生性（正規分布）

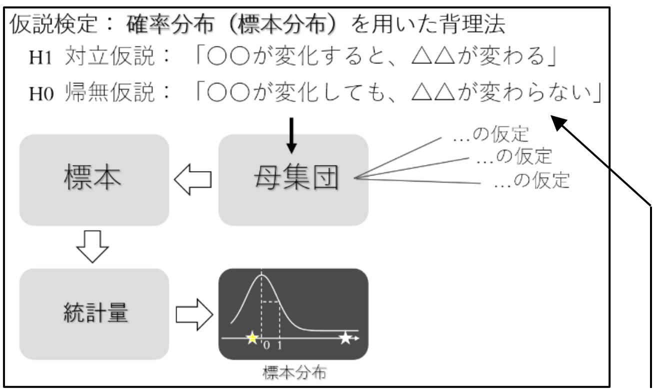
$$y_0 \sim N(\mu_0, \sigma_0^2)$$ 現代文の点
$$y_1 \sim N(\mu_1, \sigma_1^2)$$ 数学の点
$$\vdots$$
$$y_n \sim N(\mu_n, \sigma_n^2)$$ 世界史の点

+)

$$y_0 + y_1 + \cdots + y_n$$ 総合点
$$\sim N(\mu_0 + \mu_1 + \cdots + \mu_n, \sigma_0^2 + \sigma_1^2 + \cdots + \sigma_n^2)$$



ねらい



「従属変数（△△）に変化があるか」という H0 を吟味するために中心極限定理と再生性を使って$\mu_1 - \mu_0 = 0$という帰無仮説を検定。

Step 2 標準化（z 値を考える）

$$\bar{y}_1 - \bar{y}_0 \Leftarrow N\left(\mu_1 - \mu_0,\ \frac{\sigma_0^2}{n_0} + \frac{\sigma_1^2}{n_1}\right)$$

↓標準化

$$\frac{(\bar{y}_1-\bar{y}_0)-(\mu_1-\mu_0)}{\sqrt{\frac{\sigma_0^2}{n_0}+\frac{\sigma_1^2}{n_1}}} \Leftarrow N(0,1)$$

Step 3 帰無仮説を導入

$$\frac{(\bar{y}_1-\bar{y}_0)-(\mu_1-\mu_0)}{\sqrt{\frac{\sigma_0^2}{n_0}+\frac{\sigma_1^2}{n_1}}} \Leftarrow N(0,1)$$

↓帰無仮説（$\mu_1 - \mu_0 = 0$）

$$\frac{(\bar{y}_1-\bar{y}_0)}{\sqrt{\frac{\sigma_0^2}{n_0}+\frac{\sigma_1^2}{n_1}}} \Leftarrow N(0,1)$$

Step 4 補正された標準化（t 値を考える）

$$\frac{(\bar{y}_1-\bar{y}_0)}{\sqrt{\frac{\sigma_0^2}{n_0}+\frac{\sigma_1^2}{n_1}}} \Leftarrow N(0,1)$$

↓点推定

$$\frac{\bar{y}_1-\bar{y}_0}{\sqrt{\frac{\hat{\sigma}_0^2}{n_0}+\frac{\hat{\sigma}_1^2}{n_1}}} \Leftarrow \mathrm{t}\left(\frac{\left(\frac{\hat{\sigma}_0^2}{n_0}+\frac{\hat{\sigma}_1^2}{n_1}\right)^2}{\frac{\hat{\sigma}_0^4}{n_0^2(n_0-1)}+\frac{\hat{\sigma}_1^4}{n_1^2(n_1-1)}}\right)$$

⇒ Welch の t 検定

↓分散等質性の仮定（$\sigma^2 = \sigma_0^2 = \sigma_1^2$）

$$t = \frac{\bar{y}_1-\bar{y}_0}{\hat{\sigma}\sqrt{\frac{1}{n_0}+\frac{1}{n_1}}} \Leftarrow \mathrm{t}(n_0 + n_1 - 2)$$

自由度

今回の t 値である下記の量が$t(n_0+n_1-2)$に従うことの証明

$$t=\frac{(\bar{y}_1-\bar{y}_0)-(\mu_1-\mu_0)}{\hat{\sigma}\sqrt{\frac{1}{n_0}+\frac{1}{n_1}}}$$

【証明】次の F 値が自由度$1, n_0+n_1-2$の F 分布に従うことを示す

$$F=t^2 \quad =\left(\frac{(\bar{y}_1-\bar{y}_0)-(\mu_1-\mu_0)}{\hat{\sigma}\sqrt{\frac{1}{n_0}+\frac{1}{n_1}}}\right)^2$$

$$=\left(\frac{(\bar{y}_1-\bar{y}_0)-(\mu_1-\mu_0)}{\sigma/\sqrt{\frac{1}{n_0}+\frac{1}{n_1}}}\right)^2\Bigg/\left(\frac{\hat{\sigma}}{\sigma}\right)^2$$

$$(\text{分子}) \quad =\left(\frac{(\bar{y}_1-\bar{y}_0)-(\mu_1-\mu_0)}{\sigma/\sqrt{\frac{1}{n_0}+\frac{1}{n_1}}}\right)^2 \sim \chi^2(1)$$

$$=\{\text{自由度 1 の}\chi^2\text{分布に従う変数}\}/1$$

$$(\text{分母}) \quad =\frac{\hat{\sigma}^2}{\sigma^2}$$

$$=\frac{\frac{1}{n_0+n_1-2}\left\{\sum_{i=1}^{n_0}(y_{i0}-\bar{y}_0)^2+\sum_{i=1}^{n_1}(y_{i1}-\bar{y}_1)^2\right\}}{\sigma^2}$$

$$=\frac{\sum_{i=1}^{n_0}(y_{i0}-\bar{y}_0)^2+\sum_{i=1}^{n_1}(y_{i1}-\bar{y}_1)^2}{\sigma^2}\Big/(n_0+n_1-2)$$

$$=\left\{\sum_{i=1}^{n_0}\frac{(y_{i0}-\bar{y}_0)^2}{\sigma^2}+\sum_{i=1}^{n_1}\frac{(y_{i1}-\bar{y}_1)^2}{\sigma^2}\right\}\Big/(n_0+n_1-2)$$

$$=\left[\sum_{i=1}^{n_0}\frac{\{(y_{i0}-\mu_1)-(\bar{y}_0-\mu_1)\}^2}{\sigma^2}+\sum_{i=1}^{n_1}\frac{(y_{i1}-\bar{y}_1)^2}{\sigma^2}\right]\Big/(n_0+n_1-2)$$

$$=\left[\left\{\sum_{i=1}^{n_0}\frac{(y_{i0}-\mu_1)^2}{\sigma^2}-\frac{2(\bar{y}_0-\mu_1)\sum_{i=1}^{n_0}(y_{i0}-\mu_0)}{\sigma^2}+n\frac{(\bar{y}_0-\mu_0)^2}{\sigma^2}\right\}+\sum_{i=1}^{n_1}\frac{(y_{i1}-\bar{y}_1)^2}{\sigma^2}\right]\Big/(n_0+n_1-2)$$

$$=\left[\left\{\sum_{i=1}^{n_0}\left(\frac{y_{i0}-\mu_0}{\sigma}\right)^2-\frac{2n_0(\bar{y}_0-\mu_0)^2}{\sigma^2}+\frac{n_0(\bar{y}_0-\mu_0)^2}{\sigma^2}\right\}+\sum_{i=1}^{n_1}\frac{(y_{i1}-\bar{y}_1)^2}{\sigma^2}\right]\Big/(n_0+n_1-2)$$

$$=\left[\sum_{i=1}^{n_0}\left(\frac{y_{i0}-\mu_0}{\sigma}\right)^2-\left(\frac{\bar{y}_0-\mu_0}{\sigma/\sqrt{n_0}}\right)^2+\sum_{i=1}^{n_1}\left(\frac{y_{i1}-\mu_1}{\sigma}\right)^2-\left(\frac{\bar{y}_1-\mu_1}{\sigma/\sqrt{n_1}}\right)^2\right]\Big/(n_0+n_1-2)$$

$$=\{\text{自由度}n_0-1+n_1-1\text{ の}\chi^2\text{分布に従う変数}\}/(n_0+n_1-2)$$

$$\therefore\ F_i=t_i^2=\frac{\{\text{自由度 1 の}\chi^2\text{分布に従う変数}\}/1}{\{\text{自由度}n_0-1+n_1-1\text{ の}\chi^2\text{分布に従う変数}\}/(n_0+n_1-2)}$$

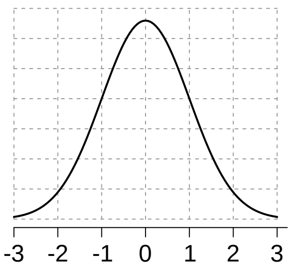
$$\sim F(1,n_0+n_1-2)$$

📖 補足 標本分布２：標準正規分布からカイ二乗分布へ

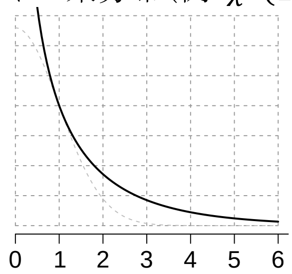
（１）カイ二乗分布 Chi-squared (Chi-square) distribution

これは、平均が 0、分散が 1 の正規分布$N(0,1)$に独立にしたがう n 個の変数（z 値）の平方和がしたがう確率分布。

標準正規分布 $N(0,1)$



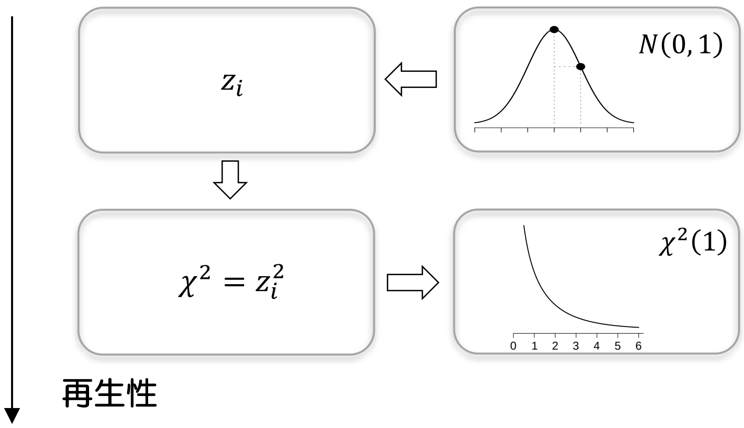
カイ二乗分布（例：$\chi^2(1)$）



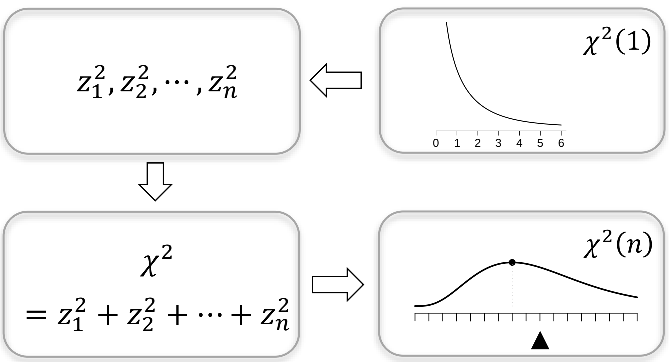
（２）カイ二乗統計量

これは、平均が 0、分散が 1 の正規分布$N(0,1)$に独立にしたがう n 個の変数（z 値）の平方和。

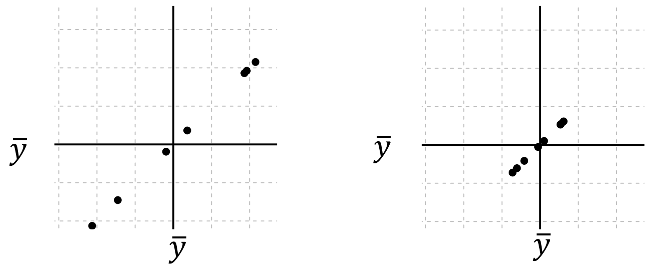
■ 自由度 1 のカイ二乗分布



■ 自由度 n のカイ二乗分布



平方和と分散



① 平方和

$$SS = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

② 分散

$$s^2 = \frac{1}{n} \sum_{i=1}^{n} (y_i - \bar{y})^2$$

③ 不偏分散

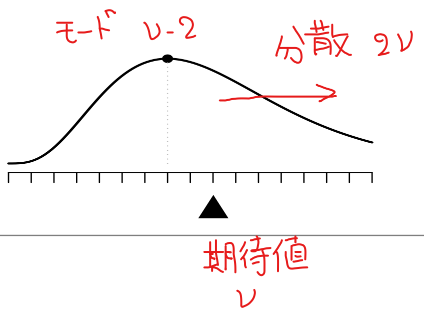
$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \bar{y})^2$$

再生性と自由度（カイ二乗分布）

$$\begin{aligned} z_1^2 &\sim \chi^2(1) \\ z_2^2 &\sim \chi^2(1) \\ &\vdots \\ +) \quad z_n^2 &\sim \chi^2(1) \\ \hline z_1^2 + z_2^2 + \cdots + z_n^2 &\sim \chi^2(n) \end{aligned}$$

カイ二乗分布の期待値とモード

自由度 ν のカイ二乗分布



期待値：	ν
分散：	2ν
モード：	$\nu - 2$

全知全能性とカイ二乗値

① 全知全能の視点

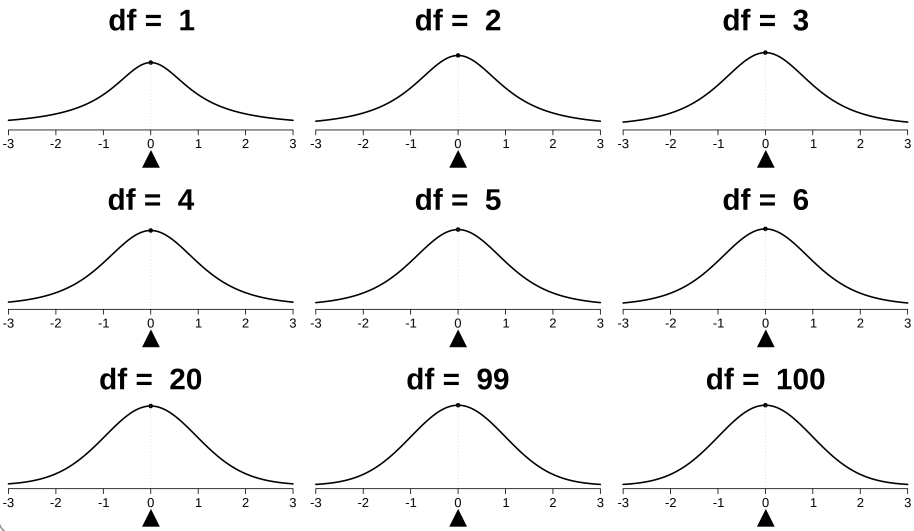
$$z_i \quad = \quad \frac{y_i - \mu}{\sigma} \qquad \sim N(0, 1)$$

なりかけ

② 非全知全能の視点

$$t_i \quad = \quad \frac{y_i - \mu}{\hat{\sigma}} \qquad \sim t(n-1)$$

t 分布と自由度



不偏分散のしたがう標本分布

$$\chi^2 = \sum_{i=1}^{n} z_i^{*2} = \sum_{i=1}^{n} \left(\frac{y_i - \bar{y}}{\sigma}\right)^2 \qquad \sim \chi^2(n-1)$$

$$= \frac{1}{\sigma^2} \sum_{i=1}^{n} (y_i - \bar{y})^2$$

$$= \frac{1}{\sigma^2}(n-1) \times \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \bar{y})^2$$

$$= \frac{(n-1)s^2}{\sigma^2} \qquad \sim \chi^2(n-1)$$

全知全能性とカイ二乗値

① 全知全能の視点

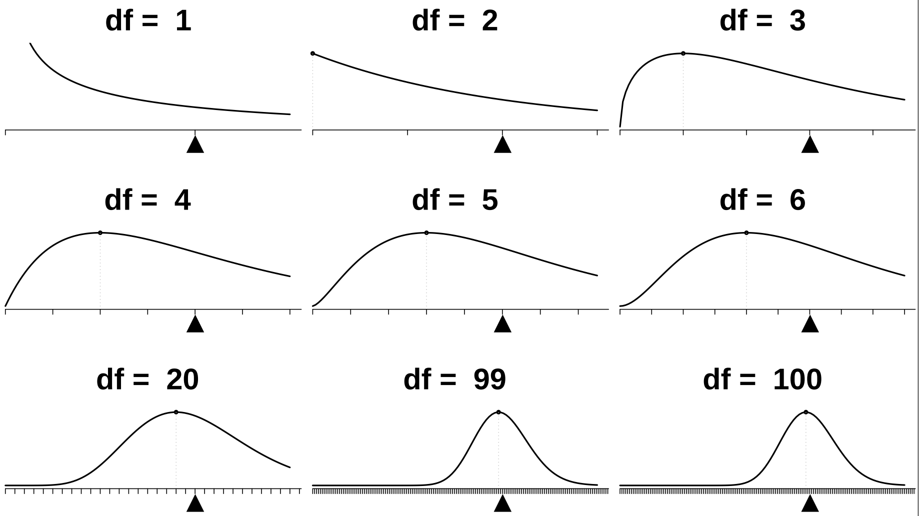
$$\chi^2 = \sum_{i=1}^{n} z_i^2 = \sum_{i=1}^{n} \left(\frac{y_i - \mu}{\sigma}\right)^2 \sim \chi^2(n)$$

なりかけ

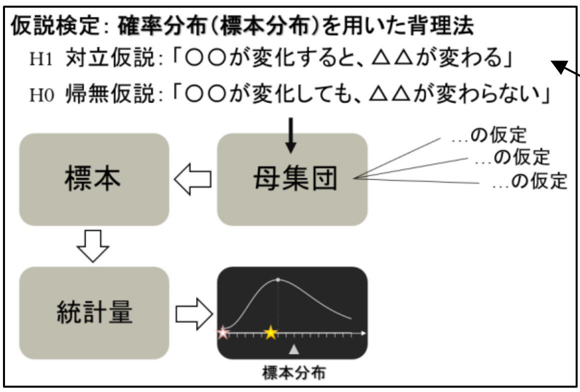
② 非全知全能の視点

$$\chi^2 = \sum_{i=1}^{n} z_i^{*2} = \sum_{i=1}^{n} \left(\frac{y_i - \bar{y}}{\sigma}\right)^2 \sim \chi^2(n-1)$$

カイ二乗分布と自由度



ねらい



平均ではなく、分散（平方和）に関する帰無仮説を立てる

Ⓟ 変数を比較する視点

(視点1) 引き算

「平均値の差」など

$\bar{y}_1 - \bar{y}_0$

↑ 再生性

(視点2) 割り算

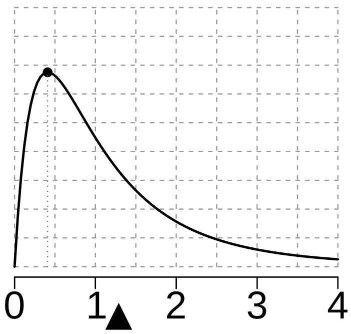
「分散の比」など

$\dfrac{s_0^2}{s_1^2}$ ← 正 ← 正

📖 補足 標本分布３：カイ二乗分布からＦ分布へ

（１） F 分布 F distribution

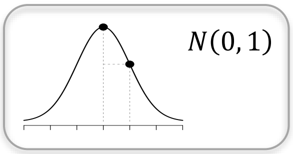
これは、二つの標準正規分布にしたがう変数の分散の比がしたがう確率分布。



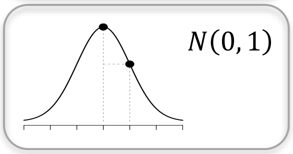
（２） F 値 F value

これは、比（割り算）の形で、標準正規分布に従う二つのグループから得られた二つの分散の違いを表した統計量。

$$z_{0i} = \frac{y_{0i} - \mu_0}{\sigma_0}$$ ⇨

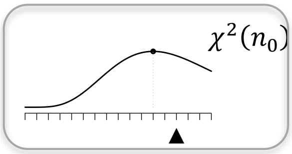


$$z_{1i} = \frac{y_{1i} - \mu_1}{\sigma_1}$$ ⇨



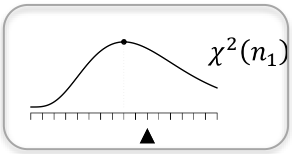
⇩ ① 再生性

$$\chi_0^2 = \Sigma_{i=1}^{n_0} z_{0i}^2$$ ⇨



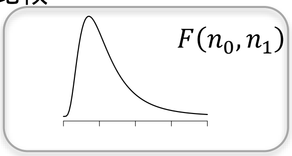
÷

$$\chi_1^2 = \Sigma_{i=1}^{n_1} z_{1i}^2$$ ⇨



⇩ ② 比による比較

$$F = \frac{\frac{1}{n_0}\chi_0^2}{\frac{1}{n_1}\chi_1^2}$$ ⇨



全知全能性とＦ値

① 全知全能の視点

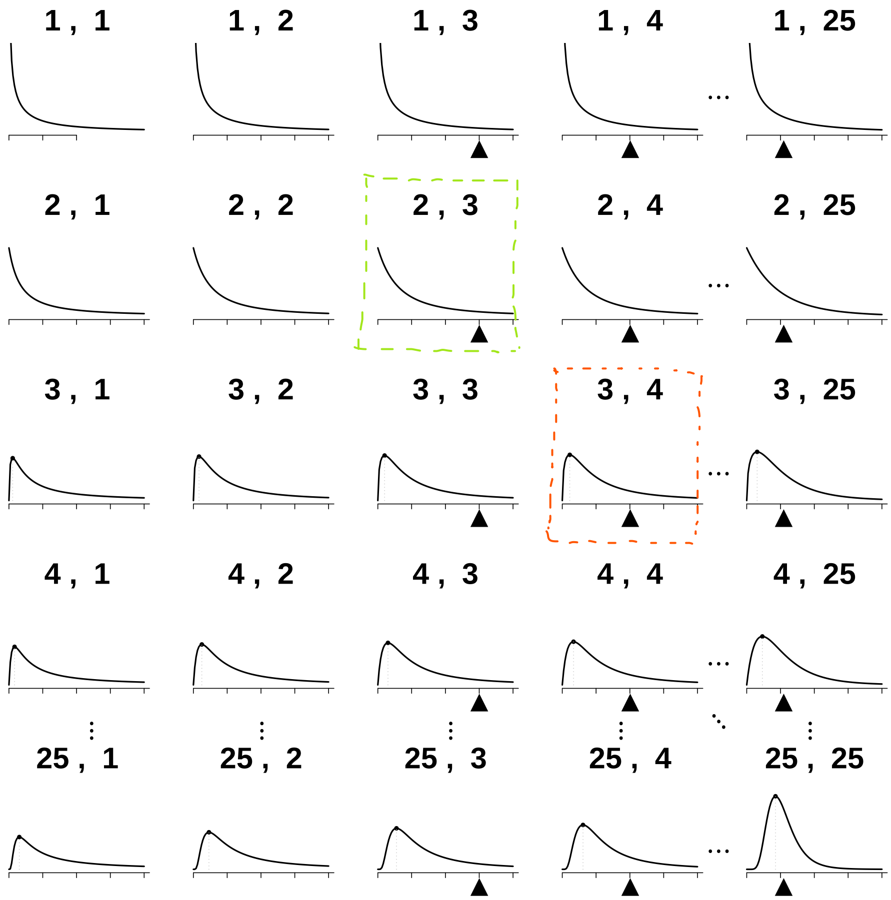
$$F = \frac{\frac{1}{n_0}\sum_{i=1}^{n_0}\left(\frac{y_{0i}-\mu_0}{\sigma_0}\right)^2}{\frac{1}{n_1}\sum_{i=1}^{n_1}\left(\frac{y_{1i}-\mu_1}{\sigma_1}\right)^2} \quad \sim F(\,n_0,\, n_1\,)$$

なりかけ

② 非全知全能の視点

$$F = \frac{\frac{1}{n_0-1}\sum_{i=1}^{n_0}\left(\frac{y_{0i}-\bar{y}_0}{\sigma_0}\right)^2}{\frac{1}{n_1-1}\sum_{i=1}^{n_1}\left(\frac{y_{1i}-\bar{y}_1}{\sigma_1}\right)^2} \quad \sim F(\,n_0-1,\, n_1-1\,)$$

Ｆ分布と自由度



等分散性が成り立つときの標本分布

「H0: 分散等質性の仮定（$\sigma^2 = \sigma_0^2 = \sigma_1^2$）」のもとでは、

① 全知全能の視点

$$F = \frac{\frac{1}{n_0}\sum_{i=1}^{n_0}\left(\frac{y_{0i}-\mu_0}{\sigma}\right)^2}{\frac{1}{n_1}\sum_{i=1}^{n_1}\left(\frac{y_{1i}-\mu_1}{\sigma}\right)^2} \sim F(n_0, n_1)$$

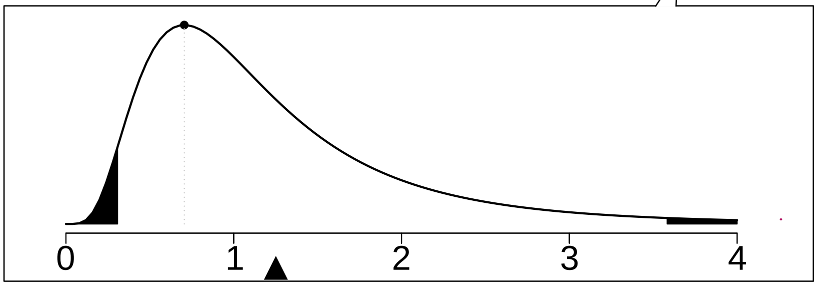
$$= \frac{\frac{1}{n_0}\sum_{i=1}^{n_0}(y_{0i}-\mu_0)^2}{\frac{1}{n_1}\sum_{i=1}^{n_1}(y_{1i}-\mu_1)^2} \sim F(n_0, n_1)$$

② 非全知全能の視点

$$F = \frac{\frac{1}{n_0-1}\sum_{i=1}^{n_0}\left(\frac{y_{0i}-\bar{y}_0}{\sigma}\right)^2}{\frac{1}{n_1-1}\sum_{i=1}^{n_1}\left(\frac{y_{1i}-\bar{y}_1}{\sigma}\right)^2} \sim F(n_0-1, n_1-1)$$

$$= \frac{\frac{1}{n_0-1}\sum_{i=1}^{n_0}(y_{0i}-\bar{y}_0)^2}{\frac{1}{n_1-1}\sum_{i=1}^{n_1}(y_{1i}-\bar{y}_1)^2}$$

$$= \frac{s_0^2}{s_1^2} \sim F(n_0-1, n_1-1)$$



Ⓟ t値とF値の関係

$$t_i = \frac{y_i - \mu}{\hat{\sigma}}$$

$$t_i^2 = \frac{(y_i - \mu)^2}{\hat{\sigma}^2}$$

$$= \frac{\frac{1}{1}\left(\frac{y_i - \mu}{\sigma}\right)^2}{\frac{\hat{\sigma}^2}{\sigma^2}}$$

$$= \frac{\frac{1}{1}\left(\frac{y_i - \mu}{\sigma}\right)^2}{\frac{1}{(n-1)}\frac{(n-1)\hat{\sigma}^2}{\sigma^2}}$$

$$= \frac{\frac{1}{1}\left(\text{自由度1の}\chi^2\text{分布に従う}\right)}{\frac{1}{n-1}\left(\text{自由度}n-1\text{の}\chi^2\text{分布に従う}\right)}$$

$$\sim F(1, n-1)$$

$$t_i^2 = F$$