

数理統計学2022目次

1 目標

▼ 2 資料

[2.1 01 Bernoulli試行と関連確率分布](#)

[2.2 02 Gauss積分, ガンマ関数, ベータ関数](#)

[2.3 03 確率分布達の解釈](#)

[2.4 04 標本分布について](#)

[2.5 05 大数の法則と中心極限定理](#)

[2.6 06 条件付き確率分布, 尤度, 推定, 記述統計](#)

▼ 2.7 07 例

[2.7.1 07-1 例: ベータ関数と二項分布の関係とその応用](#)

[2.7.2 07-2 Julia言語を使った統計学の勉強の仕方](#)

[2.8 08 P値と仮説検定と信頼区間\(1\) 一般論](#)

[2.9 09 P値と仮説検定と信頼区間\(2\) 母比率](#)

[2.10 10 P値と仮説検定と信頼区間\(3\) 母平均](#)

▼ 2.11 11 P値と仮説検定と信頼区間(4) 母比率の違い

[2.11.1 「P値と仮説検定と信頼区間\(4\) 母比率の違い」のノートブックの内容をJulia言語で直接実行するためのヒント](#)

[2.12 12 P値と仮説検定と信頼区間\(5\) 母平均の差](#)

[2.13 13 誤用を避けるための注意](#)

[2.14 14 回帰](#)

[2.15 15 まとめ](#)

▼ 3 参考資料

[3.1 以上で解説できなかったことに関するノート](#)

数理統計学2022

すべてをまとめたPDF:

- [Statistics2022.pdf](#) (<https://github.com/genkuroki/Statistics/blob/master/2022/Statistics2022.pdf>) [[download](#) (<https://github.com/genkuroki/Statistics/raw/master/2022/Statistics2022.pdf>)]
- [Statistics2022handwritten.pdf](#) (<https://github.com/genkuroki/Statistics/blob/master/2022/handwritten/Statistics2022handwritten.pdf>) [[download](#) (<https://github.com/genkuroki/Statistics/raw/master/2022/handwritten/Statistics2022handwritten.pdf>)]

これはおまけで作っているPDFファイルなので更新は遅れる予定. 以下の個別のPDFファイルの方が常に新しいバージョンになっている. ファイル名に2022とあるので2022年の資料だと思う人もいるかもしれないが, 2023年以降も少しずつ書き変えている.

このファイルのipynb版とpdf版:

- [Jupyter notebook](#) (<https://github.com/genkuroki/Statistics/blob/master/2022/00-0%20README.ipynb>) [[nbviewer](#) (<https://nbviewer.org/github/genkuroki/Statistics/blob/master/2022/00-0%20README.ipynb>)]
- [pdf](#) (<https://github.com/genkuroki/Statistics/blob/master/2022/00-0%20README.pdf>) [[download](#) (<https://github.com/genkuroki/Statistics/raw/master/2022/00-0%20README.pdf>)]

記号法については次のメモも参照せよ:

- [Memo.md](#) (<https://github.com/genkuroki/Statistics/blob/master/2022/Memo.md>)
- [Jupyter notebook](#) (<https://github.com/genkuroki/Statistics/blob/master/2022/00-1%20Memo.ipynb>) [[nbviewer](#) (<https://nbviewer.org/github/genkuroki/Statistics/blob/master/2022/00-1%20Memo.ipynb>)]
- [pdf](#) (<https://github.com/genkuroki/Statistics/blob/master/2022/00-1%20Memo.pdf>) [[download](#) (<https://github.com/genkuroki/Statistics/raw/master/2022/00-1%20Memo.pdf>)]

1 目標

- 確率変数の扱い方
- 大数の法則と中心極限定理
- よく使われる確率分布達に関する知識
- 尤度と推定と要約統計
- 検定と信頼区間を表裏一体のものとして理解すること
- P値関数について理解すること
- 簡単な場合について具体的な数値の計算を行えるようになること(重要!)

2 資料

[Jupyter](https://jupyter.org/) (<https://jupyter.org/>) notebook をオンラインで読むなら [nbviewer](https://nbviewer.org/) (<https://nbviewer.org/>) 経由の閲覧がおすすめです。オフラインでも読めるようにするためには pdf ファイルをダウンロードしてください。

[Julia言語](https://julialang.org/) (<https://julialang.org/>)を使って数値の計算やグラフの作画を行っています。必要最小限の使い方については次の資料を見て下さい:

- 07-2 Julia言語を使った統計学の勉強の仕方 (<https://nbviewer.org/github/genkuroki/Statistics/blob/master/2022/07-2%20How%20to%20use%20Julia%20language%20for%20learning%20statistics.ipynb>)

[WolframAlpha](https://www.wolframalpha.com/) (<https://www.wolframalpha.com/>)も併用しています。

明らかな誤りをまだ大量に含んでいるものと思われます。ごめんなさい。適切に訂正しながら読んでください。随時訂正更新しています。

2.1 01 Bernoulli試行と関連確率分布

- [Jupyter notebook](https://github.com/genkuroki/Statistics/blob/master/2022/01%20Bernoulli%20trial%20and%20related%20distributions.ipynb)
(<https://github.com/genkuroki/Statistics/blob/master/2022/01%20Bernoulli%20trial%20and%20related%20distributions.ipynb>)
[\[nbviewer\]](https://nbviewer.org/github/genkuroki/Statistics/blob/master/2022/01%20Bernoulli%20trial%20and%20related%20distributions)
(<https://nbviewer.org/github/genkuroki/Statistics/blob/master/2022/01%20Bernoulli%20trial%20and%20related%20distributions>)
- [pdf](https://github.com/genkuroki/Statistics/blob/master/2022/01%20Bernoulli%20trial%20and%20related%20distributions.pdf)
(<https://github.com/genkuroki/Statistics/blob/master/2022/01%20Bernoulli%20trial%20and%20related%20distributions.pdf>)
[\[download\]](https://github.com/genkuroki/Statistics/blob/master/2022/01%20Bernoulli%20trial%20and%20related%20distributions.pdf)
(<https://github.com/genkuroki/Statistics/raw/master/2022/01%20Bernoulli%20trial%20and%20related%20distributions.pdf>)
- [hand-written note](https://github.com/genkuroki/Statistics/blob/master/2022/handwritten/%E6%95%B0%E7%90%86%E7%B5%B1%E8%A8%88%9C)
(<https://github.com/genkuroki/Statistics/blob/master/2022/handwritten/%E6%95%B0%E7%90%86%E7%B5%B1%E8%A8%88%9C>)
[\[download\]](https://github.com/genkuroki/Statistics/blob/master/2022/handwritten/%E6%95%B0%E7%90%86%E7%B5%B1%E8%A8%88%9C)
(<https://github.com/genkuroki/Statistics/raw/master/2022/handwritten/%E6%95%B0%E7%90%86%E7%B5%B1%E8%A8%88%9C>)
- [hand-written note](https://github.com/genkuroki/Statistics/blob/master/2022/handwritten/%E6%95%B0%E7%90%86%E7%B5%B1%E8%A8%88%9C)
(<https://github.com/genkuroki/Statistics/blob/master/2022/handwritten/%E6%95%B0%E7%90%86%E7%B5%B1%E8%A8%88%9C>)
[\[download\]](https://github.com/genkuroki/Statistics/blob/master/2022/handwritten/%E6%95%B0%E7%90%86%E7%B5%B1%E8%A8%88%9C)
(<https://github.com/genkuroki/Statistics/raw/master/2022/handwritten/%E6%95%B0%E7%90%86%E7%B5%B1%E8%A8%88%9C>)

定義された確率分布

- カテゴリカル分布**: $\text{Categorical}(p_1, \dots, p_n)$
- 成功確率 p の **Bernoulli分布**: $\text{Bernoulli}(p)$
- 試行回数 n , 成功確率 p の **Bernoulli試行の分布**: $\text{Bernoulli}(p)^n$
- 試行回数 n , 成功確率 p の **二項分布**: $\text{Binomial}(n, p)$ (n 回のBernoulli試行での成功回数の分布)
- 成功回数 k , 成功確率 p の **負の二項分布**: $\text{NegativeBinomial}(n, p) = \text{NegBin}(n, p)$ (Bernoulli試行をちょうど k 回成功するまで続けたときの失敗の回数の分布)
- 成功確率 p の **幾何分布**: $\text{Geometric}(p) = \text{NegativeBinomial}(1, p)$ (Bernoulli試行を成功するまで続けたときの失敗の回数の分布)
- 0から1のあいだの **一様分布**: $\text{Uniform}(0, 1)$ (コンピュータでの一様乱数関数 `rand()` のモデル化)
- 試行回数 n の **一様乱数生成の繰り返しの分布**: $\text{Uniform}(0, 1)^n$
- 平均 μ , 分散 σ^2 (標準偏差 σ) の **正規分布**: $\text{Normal}(\mu, \sigma)$
- 標準正規分布**: $\text{Normal}() = \text{Normal}(0, 1)$
- 分布 D の **アフィン変換**: $aD + b$
- 分布 D の **スケール変換**: aD ($a > 0$)
- 分布 D の **平行移動**: $D + b$
- 分布 D の **逆数**: $1/D$

2.2 02 Gauss積分, ガンマ関数, ベータ関数

- [Jupyter notebook](https://github.com/genkuroki/Statistics/blob/master/2022/02%20Gaussian%20integrals%2C%20Gamma%20and%20Beta%20f)
(<https://github.com/genkuroki/Statistics/blob/master/2022/02%20Gaussian%20integrals%2C%20Gamma%20and%20Beta%20f>)
[\[nbviewer\]](https://nbviewer.org/github/genkuroki/Statistics/blob/master/2022/02%20Gaussian%20integrals%2C%20Gamma%20and%20Beta%20f)
(<https://nbviewer.org/github/genkuroki/Statistics/blob/master/2022/02%20Gaussian%20integrals%2C%20Gamma%20and%20Beta%20f>)
- [pdf](https://github.com/genkuroki/Statistics/blob/master/2022/02%20Gaussian%20integrals%2C%20Gamma%20and%20Beta%20f)
(<https://github.com/genkuroki/Statistics/blob/master/2022/02%20Gaussian%20integrals%2C%20Gamma%20and%20Beta%20f>)
[\[download\]](https://github.com/genkuroki/Statistics/blob/master/2022/02%20Gaussian%20integrals%2C%20Gamma%20and%20Beta%20f)
(<https://github.com/genkuroki/Statistics/raw/master/2022/02%20Gaussian%20integrals%2C%20Gamma%20and%20Beta%20f>)
- [hand-written note](https://github.com/genkuroki/Statistics/blob/master/2022/handwritten/%E6%95%B0%E7%90%86%E7%B5%B1%E8%A8%88%9C)
(<https://github.com/genkuroki/Statistics/blob/master/2022/handwritten/%E6%95%B0%E7%90%86%E7%B5%B1%E8%A8%88%9C>)
[\[download\]](https://github.com/genkuroki/Statistics/blob/master/2022/handwritten/%E6%95%B0%E7%90%86%E7%B5%B1%E8%A8%88%9C)
(<https://github.com/genkuroki/Statistics/raw/master/2022/handwritten/%E6%95%B0%E7%90%86%E7%B5%B1%E8%A8%88%9C>)
- [hand-written note](https://github.com/genkuroki/Statistics/blob/master/2022/handwritten/%E6%95%B0%E7%90%86%E7%B5%B1%E8%A8%88%9C)
(<https://github.com/genkuroki/Statistics/blob/master/2022/handwritten/%E6%95%B0%E7%90%86%E7%B5%B1%E8%A8%88%9C>)
[\[download\]](https://github.com/genkuroki/Statistics/blob/master/2022/handwritten/%E6%95%B0%E7%90%86%E7%B5%B1%E8%A8%88%9C)
(<https://github.com/genkuroki/Statistics/raw/master/2022/handwritten/%E6%95%B0%E7%90%86%E7%B5%B1%E8%A8%88%9C>)

[\(https://github.com/genkuroki/Statistics/raw/master/2022/handwritten/%E6%95%B0%E7%90%86%E7%B5%B1%E8%A8%88%](https://github.com/genkuroki/Statistics/raw/master/2022/handwritten/%E6%95%B0%E7%90%86%E7%B5%B1%E8%A8%88%)

- 正規分布 (再): $\text{Normal}(\mu, \sigma) = \mu + \sigma \text{Normal}(0, 1)$
- 形状 α , スケール θ の **ガンマ分布**: $\text{Gamma}(\alpha, \theta) = \theta \text{Gamma}(\alpha, 1)$ (平均的なイベント発生間隔が θ のときにちょうど α 回イベントが生じるまでにかかる時間の分布)
- 平均 θ の **指数分布**: $\text{Exponential}(\theta) = \text{Gamma}(1, \theta)$ (平均的なイベント発生間隔が θ のときにイベントが発生するまでにかかる時間の分布)
- 中央値 a , スケール b の **Laplace分布**: $\text{Laplace}(a, b)$
- 形状 α , スケール θ の **逆ガンマ分布**: $\text{InverseGamma}(\alpha, \theta) = 1/\text{Gamma}(\alpha, 1/\theta)$
- 自由度 ν の **χ^2 分布**(カイ二乗分布): $\text{Chisq}(\nu) = \text{Gamma}(\nu/2, 2)$
- 形状 α , スケール θ の **Weibull分布**: $\text{Weibull}(\alpha, \theta)$
- パラメータ α, β の **ベータ分布**: $\text{Beta}(\alpha, \beta)$ ($\alpha + \beta - 1$ 個の一樣乱数の中で α 番目に小さな (= β 番目に大きな) 値の分布)
- 自由度 ν の **t分布**: $\text{TDist}(\nu)$ (平均 0 の正規分布で分散の逆数が $\text{Chisq}(\nu)/\nu$ に従う場合, $\text{BetaPrime}(1/2, \nu/2)$ と密接に関係)
- パラメータ α, β の **ベータプライム分布**: $\text{BetaPrime}(\alpha, \beta)$ (確率 t に関するベータ分布 $\text{Beta}(\alpha, \beta)$ をオッズ $u = t/(1 - t) = 1/(1 - t) - 1$ に関する分布に変換したもの)
- 自由度 ν_1, ν_2 の **F分布**: $\text{FDist}(\nu_1, \nu_2) = (\nu_2/\nu_1) \text{BetaPrime}(\nu_1/2, \nu_2/2)$
- Dirichlet分布**: $\text{Dirichlet}(\alpha_1, \dots, \alpha_n)$ ($n = 2$ の場合がベータ分布と一致)

- [Jupyter notebook](https://github.com/genkuroki/Statistics/blob/master/2022/03%20Interpretation%20of%20probability%20distributions.ipynb)
(<https://github.com/genkuroki/Statistics/blob/master/2022/03%20Interpretation%20of%20probability%20distributions.ipynb>)
[\[nbviewer\]](#)
(<https://nbviewer.org/github/genkuroki/Statistics/blob/master/2022/03%20Interpretation%20of%20probability%20distributions.ip>)
- [pdf](https://github.com/genkuroki/Statistics/blob/master/2022/03%20Interpretation%20of%20probability%20distributions.pdf) (<https://github.com/genkuroki/Statistics/blob/master/2022/03%20Interpretation%20of%20probability%20distributions.pdf>)
[\[download\]](#)
(<https://github.com/genkuroki/Statistics/raw/master/2022/03%20Interpretation%20of%20probability%20distributions.pdf>)
- [hand-written note](https://github.com/genkuroki/Statistics/blob/master/2022/handwritten/%E6%95%B0%E7%90%86%E7%B5%B1%E8%A8%88%9A%E6%95%B0%E7%90%86%E7%B5%B1%E8%A8%88%9A.pdf)
(<https://github.com/genkuroki/Statistics/blob/master/2022/handwritten/%E6%95%B0%E7%90%86%E7%B5%B1%E8%A8%88%9A%E6%95%B0%E7%90%86%E7%B5%B1%E8%A8%88%9A.pdf>)
[\[download\]](#)
(<https://github.com/genkuroki/Statistics/raw/master/2022/handwritten/%E6%95%B0%E7%90%86%E7%B5%B1%E8%A8%88%9A%E6%95%B0%E7%90%86%E7%B5%B1%E8%A8%88%9A.pdf>)

- ・ **対数正規分布**: $\text{LogNormal}(\mu, \sigma) = \exp(\text{Normal}(\mu, \sigma))$
- ・ 平均 λ の **Poisson分布**(ポアソン分布): $\text{Poisson}(\lambda)$
- ・ **負の二項分布**: $\text{NegativeBinomial}(\alpha, 1/(1 + \theta))$ (Poisson分布の平均パラメータ λ がガンマ分布に従う場合)
- ・ a から b のあいだの **一様分布**: $\text{Uniform}(a, b)$
- ・ **ベータ二項分布**: $\text{BetaBinomial}(n, \alpha, \beta)$ (二項分布の成功確率パラメータがベータ分布に従う場合)
- ・ **超幾何分布**: $\text{Hypergeometric}(s, f, n)$ (非復元抽出の分布)
- ・ **ベータ負の二項分布**: $\text{NegativeBetaBinomial}(n, \alpha, \beta)$ (負の二項分布の成功確率パラメータがベータ分布に従う場合)
- ・ 試行回数 n の **Pólyaの壺の確率分布**: $\text{Pólya}(n, \alpha, \beta)$ (Bernoulli試行の分布で成功確率パラメータがベータ分布に従う場合)

- Jupyter notebook (<https://github.com/genkuroki/Statistics/blob/master/2022/04%20Distribution%20of%20samples.ipynb>)
[nbviewer (<https://nbviewer.org/github/genkuroki/Statistics/blob/master/2022/04%20Distribution%20of%20samples.ipynb>)]
- pdf (<https://github.com/genkuroki/Statistics/blob/master/2022/04%20Distribution%20of%20samples.pdf>) [[download](#)]
(<https://github.com/genkuroki/Statistics/raw/master/2022/04%20Distribution%20of%20samples.pdf>)]
- [hand-written note](#)
[\(https://github.com/genkuroki/Statistics/blob/master/2022/handwritten/%E6%95%B0%E7%90%86%E7%B5%B1%E8%A8%88%](https://github.com/genkuroki/Statistics/blob/master/2022/handwritten/%E6%95%B0%E7%90%86%E7%B5%B1%E8%A8%88%)
[[download](#)]
[\(https://github.com/genkuroki/Statistics/raw/master/2022/handwritten/%E6%95%B0%E7%90%86%E7%B5%B1%E8%A8%88%](https://github.com/genkuroki/Statistics/raw/master/2022/handwritten/%E6%95%B0%E7%90%86%E7%B5%B1%E8%A8%88%)

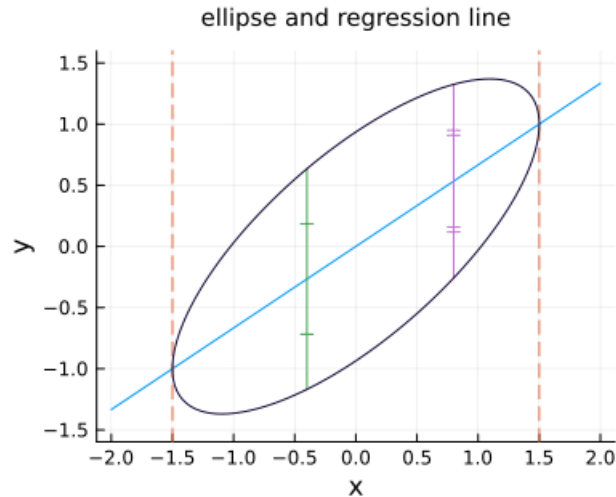
- Anscombeの例(アンスコム の例)
 - [csv file \(https://github.com/genkuroki/Statistics/blob/master/2022/data/anscombe.csv\)](https://github.com/genkuroki/Statistics/blob/master/2022/data/anscombe.csv), [raw version \(https://raw.githubusercontent.com/genkuroki/Statistics/master/2022/data/anscombe.csv\)](https://raw.githubusercontent.com/genkuroki/Statistics/master/2022/data/anscombe.csv)
 - [transposed version \(https://github.com/genkuroki/Statistics/blob/master/2022/data/anscombe_transposed.csv\)](https://github.com/genkuroki/Statistics/blob/master/2022/data/anscombe_transposed.csv), [transposed raw version \(https://raw.githubusercontent.com/genkuroki/Statistics/master/2022/data/anscombe_transposed.csv\)](https://raw.githubusercontent.com/genkuroki/Statistics/master/2022/data/anscombe_transposed.csv)

- 分布達 D_1, \dots, D_n の積: $D_1 \times \dots \times D_n$
- 分布 D の累乗: D^n

- 分布 D のサイズ n の 標本分布: D^n

作成された動画

楕円と回帰直線の関係 (https://github.com/genkuroki/Statistics/blob/master/2022/images/ellipse_and_regressionline.gif)



2.5 05 大数の法則と中心極限定理

- Jupyter notebook (<https://github.com/genkuroki/Statistics/blob/master/2022/05%20Central%20limit%20theorem.ipynb>)
[nbviewer (<https://nbviewer.org/github/genkuroki/Statistics/blob/master/2022/05%20Central%20limit%20theorem.ipynb>)]
- pdf (<https://github.com/genkuroki/Statistics/blob/master/2022/05%20Central%20limit%20theorem.pdf>) [download (<https://github.com/genkuroki/Statistics/raw/master/2022/05%20Central%20limit%20theorem.pdf>)]
- hand-written note
(<https://github.com/genkuroki/Statistics/blob/master/2022/handwritten/%E6%95%B0%E7%90%86%E7%B5%B1%E8%A8%88%E6%95%B0%E7%90%86%E7%B5%B1%E8%A8%88.pdf>)
[download (<https://github.com/genkuroki/Statistics/raw/master/2022/handwritten/%E6%95%B0%E7%90%86%E7%B5%B1%E8%A8%88%E6%95%B0%E7%90%86%E7%B5%B1%E8%A8%88.pdf>)]

大数の法則の数値例

- $(X_1 + X_2 + \dots + X_n)/n$ の分布
 - Bernoulli試行
 - 正規分布の標本分布
 - ガンマ分布 $\text{Gamma}(2, 3)$ の標本分布

大数の法則が成立しない場合の数値例

- Cuchy分布の標本分布
- Pólyaの壺 = 事前分布がベータ分布のBernoulli試行

ランダムウォークの数値例

- $X_1 + X_2 + \dots + X_n$ の分布
 - Bernoulli試行
 - 正規分布の標本分布
 - ガンマ分布 $\text{Gamma}(2, 3)$ の標本分布
 - Cuchy分布の標本分布
 - Pólyaの壺 = 事前分布がベータ分布のBernoulli試行

中心極限定理の数値例

- 対称な分布
 - 二項分布
 - 一様分布の標本分布
 - ベータ分布 $\text{Beta}(0.5, 0.5)$
 - Bernoulli試行 $p = 0.5$
 - Laplace分布の標本分布
- 非対称な分布
 - ガンマ分布 $\text{Gamma}(3, 1)$ の標本分布
 - Bernoulli試行 ($p = 0.25$) の標本分布
 - Poisson分布 (期待値1) の標本分布
 - 指数分布 (期待値1) の標本分布
 - χ^2 分布 (自由度1) の標本分布
 - 混合正規分布の標本分布
 - 対数正規分布の標本分布

正規分布で近似される確率分布の数値例

- デルタ法の応用
 - 自由度が大きな χ^2 分布の逆数
 - 自由度が大きな χ^2 分布の対数
 - 2つのパラメータが両方とも大きなベータ分布
- 適当なパラメータ領域において正規分布で近似される分布
 - 二項分布
 - 負の二項分布
 - Poisson分布
 - ガンマ分布
 - χ^2 分布
 - 逆ガンマ分布
 - ベータ分布
 - t分布
 - ベータプライム分布
 - F分布
 - 超幾何分布
 - ベータ二項分布
 - ベータ負の二項分布

2.6 06 条件付き確率分布, 尤度, 推定, 記述統計

- [Jupyter notebook](#) (<https://github.com/genkuroki/Statistics/blob/master/2022/06%20Conditional%20distribution%2C%20likelihood%2C%20estimation.ipynb>)
- [pdf](#) (<https://github.com/genkuroki/Statistics/blob/master/2022/06%20Conditional%20distribution%2C%20likelihood%2C%20estimation.pdf>)
- [hand-written note](#) (<https://github.com/genkuroki/Statistics/blob/master/2022/handwritten/%E6%95%B0%E7%90%86%E7%B5%B1%E8%A8%88%E6%95%B0%E7%90%86%E7%B5%B1%E8%A8%88.pdf>)

演習用サンプル

- [Datasaurusの例 \(データサウルスの例\)](#) (<http://www.thefunctionalart.com/2016/08/download-datasaurus-never-trust-summary.html>)
 - [csv file](#) (https://github.com/genkuroki/Statistics/blob/master/2022/data/Datasaurus_data.csv), [raw version](#) (https://raw.githubusercontent.com/genkuroki/Statistics/master/2022/data/Datasaurus_data.csv)
 - [x座標のみ](#) (https://github.com/genkuroki/Statistics/blob/master/2022/data/Datasaurus_X.txt), [y座標のみ](#) (https://github.com/genkuroki/Statistics/blob/master/2022/data/Datasaurus_Y.txt)
 - [x座標のみコンマ付き](#) (https://github.com/genkuroki/Statistics/blob/master/2022/data/Datasaurus_X_with_commas.txt), [y座標のみコンマ付き](#) (https://github.com/genkuroki/Statistics/blob/master/2022/data/Datasaurus_Y_with_commas.txt)

この例に関する詳しい解説が次の場所にある:

- Justin Matejka, George Fitzmaurice. Same Stats, Different Graphs: Generating Datasets with Varied Appearance and Identical Statistics through Simulated Annealing. Honourable Mention, ACM SIGCHI Conference on Human Factors in Computing Systems, 2017 [[link \(https://www.autodesk.com/research/publications/same-stats-different-graphs\)](https://www.autodesk.com/research/publications/same-stats-different-graphs)]

関連の情報が [datasaurus same stats](#) を検索 (<https://www.google.com/search?q=datasaurus+same+stats>) すれば得られる。

定義された確率分布

- **2変量正規分布**: $MvNormal(\mu, \Sigma)$ (**多変量正規分布** でも同じ記号を使う.)
- **2×2の分割表の分布**
 - 4つのPoisson分布の積
 - 四項分布
 - 2つの二項分布の積
 - Fisherの非心超幾何分布

2.7 07 例

2.7.1 07-1 例: ベータ関数と二項分布の関係とその応用

- [Jupyter notebook](#) (<https://github.com/genkuroki/Statistics/blob/master/2022/07-1%20Relationship%20between%20beta%20and%20binomial%20distributions.ipynb>) [[nbviewer](#)]

- (<https://nbviewer.org/github/genkuroki/Statistics/blob/master/2022/07-1%20Relationship%20between%20beta%20and%20binomial%20distributions.ipynb>)
- pdf (<https://github.com/genkuroki/Statistics/blob/master/2022/07-1%20Relationship%20between%20beta%20and%20binomial%20distributions.pdf>) [[download](#) (<https://github.com/genkuroki/Statistics/raw/master/2022/07-1%20Relationship%20between%20beta%20and%20binomial%20distributions.pdf>)]

2.7.2 07-2 Julia言語を使った統計学の勉強の仕方

- Jupyter notebook (<https://github.com/genkuroki/Statistics/blob/master/2022/07-2%20How%20to%20use%20Julia%20language%20for%20learning%20statistics.ipynb>) [[nbviewer](#) (<https://nbviewer.org/github/genkuroki/Statistics/blob/master/2022/07-2%20How%20to%20use%20Julia%20language%20for%20learning%20statistics.ipynb>)]
- pdf (<https://github.com/genkuroki/Statistics/blob/master/2022/07-2%20How%20to%20use%20Julia%20language%20for%20learning%20statistics.pdf>) [[download](#) (<https://github.com/genkuroki/Statistics/raw/master/2022/07-2%20How%20to%20use%20Julia%20language%20for%20learning%20statistics.pdf>)]

2.8 08 P値と仮説検定と信頼区間(1) 一般論

- Jupyter notebook (<https://github.com/genkuroki/Statistics/blob/master/2022/08%20Hypothesis%20testing%20and%20confidence%20interval%20%20General%20theory.ipynb>) [[nbviewer](#) (<https://nbviewer.org/github/genkuroki/Statistics/blob/master/2022/08%20Hypothesis%20testing%20and%20confidence%20interval%20%20General%20theory.ipynb>)]
- pdf (<https://github.com/genkuroki/Statistics/blob/master/2022/08%20Hypothesis%20testing%20and%20confidence%20interval%20%20General%20theory.pdf>) [[download](#) (<https://github.com/genkuroki/Statistics/raw/master/2022/08%20Hypothesis%20testing%20and%20confidence%20interval%20%20General%20theory.pdf>)]

P値関数 の概念が最も重要である。

この一連のノートでは **信頼区間は常に対応するP値関数を使って定義される** ことになる。

関連講義動画

- 京都大学大学院医学研究科 聴講コース
臨床研究者のための生物統計学「仮説検定とP値の誤解」
佐藤 俊哉 医学研究科教授
<https://youtu.be/vz9cZnB1d1c> (<https://youtu.be/vz9cZnB1d1c>)

この講義動画は「[P値に関するASA声明](https://www.biometrics.gr.jp/news/all/ASA.pdf) (<https://www.biometrics.gr.jp/news/all/ASA.pdf>)」の考え方の非常に良い解説になっている。

2.9 09 P値と仮説検定と信頼区間(2) 母比率

- Jupyter notebook (<https://github.com/genkuroki/Statistics/blob/master/2022/09%20Hypothesis%20testing%20and%20confidence%20interval%20%20Proportion.ipynb>) [[nbviewer](#) (<https://nbviewer.org/github/genkuroki/Statistics/blob/master/2022/09%20Hypothesis%20testing%20and%20confidence%20interval%20%20Proportion.ipynb>)]
- pdf (<https://github.com/genkuroki/Statistics/blob/master/2022/09%20Hypothesis%20testing%20and%20confidence%20interval%20%20Proportion.pdf>) [[download](#) (<https://github.com/genkuroki/Statistics/raw/master/2022/09%20Hypothesis%20testing%20and%20confidence%20interval%20%20Proportion.pdf>)]

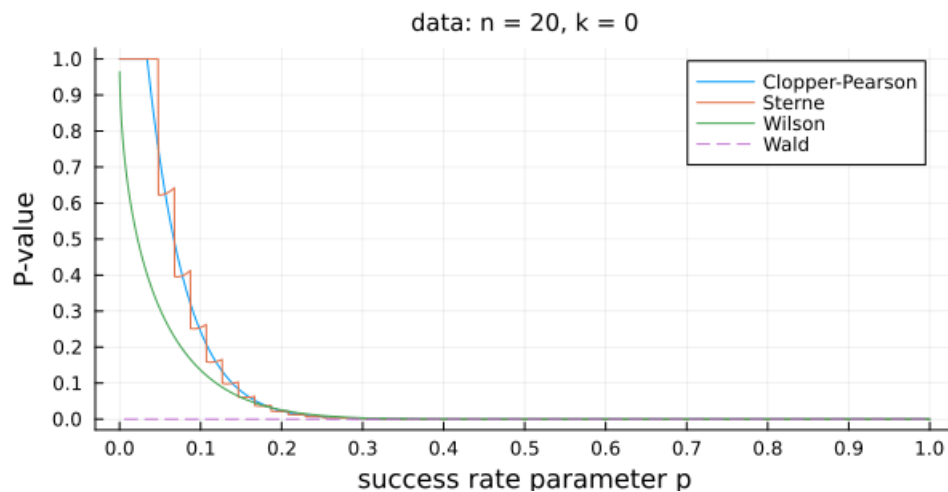
定義された信頼区間

二項検定に付随する4種の信頼区間

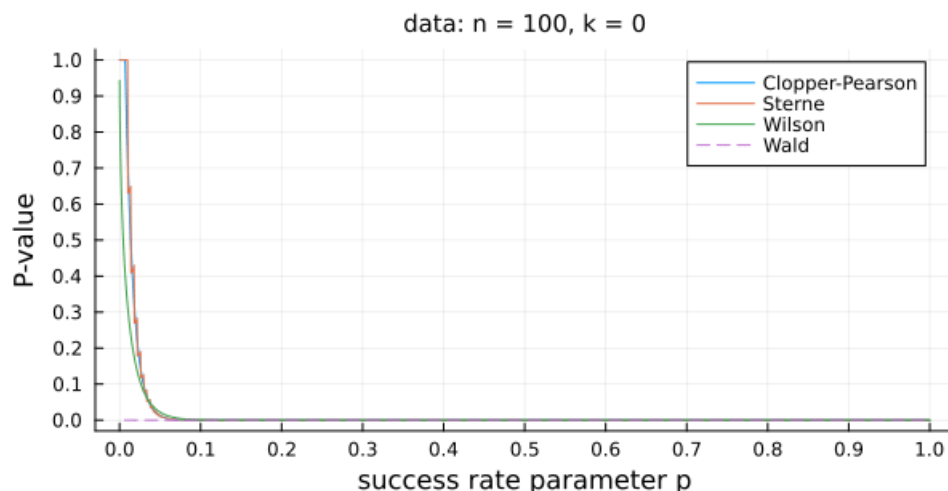
- Clopper-Pearsonの信頼区間
- Sterneの信頼区間
- Wilsonの信頼区間
- Waldの信頼区間

作成された動画

[n=20の二項分布モデルのP値関数](https://github.com/genkuroki/Statistics/blob/master/2022/images/pvaluefunction20.gif) (<https://github.com/genkuroki/Statistics/blob/master/2022/images/pvaluefunction20.gif>)



n=100の二項分布モデルのP値関数 (<https://github.com/genkuroki/Statistics/blob/master/2022/images/pvaluefunction100.gif>).



追加資料

- 比率のP値関数達
 - Jupyter notebook (<https://github.com/genkuroki/public/blob/main/0036/P-value%20functions%20of%20proportions.ipynb>) [nbviewer (<https://nbviewer.org/github/genkuroki/public/blob/main/0036/P-value%20functions%20of%20proportions.ipynb>)]
 - pdf (<https://github.com/genkuroki/public/blob/main/0036/P-value%20functions%20of%20proportions.pdf>) [download (<https://github.com/genkuroki/public/raw/main/0036/P-value%20functions%20of%20proportions.pdf>)]

2.10 10 P値と仮説検定と信頼区間(3) 母平均

- Jupyter notebook (<https://github.com/genkuroki/Statistics/blob/master/2022/10%20Hypothesis%20testing%20and%20confidence%20interval%20%20Mean.ipynb>) [nbviewer (<https://nbviewer.org/github/genkuroki/Statistics/blob/master/2022/10%20Hypothesis%20testing%20and%20confidence%20interval%20%20Mean.ipynb>)]
- pdf (<https://github.com/genkuroki/Statistics/blob/master/2022/10%20Hypothesis%20testing%20and%20confidence%20interval%20%20Mean.pdf>) [download (<https://github.com/genkuroki/Statistics/raw/master/2022/10%20Hypothesis%20testing%20and%20confidence%20interval%20%20Mean.pdf>)]

定義された信頼区間

平均の信頼区間

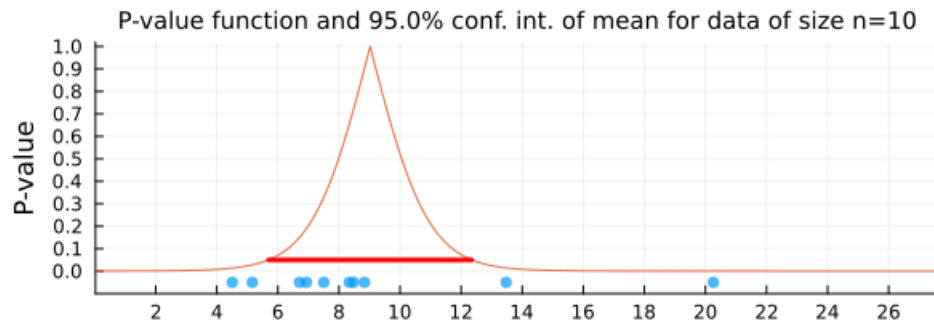
- 標準正規分布を使って計算される信頼区間
- t分布を使って計算される信頼区間

後者のみを使用する.

定義された確率分布

- 自由度 ν の χ 分布: $\text{Chi}(\nu) = \sqrt{\text{Chisq}(\nu)}$

作成された動画



2.11 11 P値と仮説検定と信頼区間(4) 母比率の違い

- [Jupyter notebook](#)
(<https://github.com/genkuroki/Statistics/blob/master/2022/11%20Hypothesis%20testing%20and%20confidence%20interval%20%20Two%20proportions.ipynb>) [[nbviewer](#)]
(<https://nbviewer.org/github/genkuroki/Statistics/blob/master/2022/11%20Hypothesis%20testing%20and%20confidence%20interval%20%20Two%20proportions.ipynb>)]
- [pdf](#)
(<https://github.com/genkuroki/Statistics/blob/master/2022/11%20Hypothesis%20testing%20and%20confidence%20interval%20%20Two%20proportions.pdf>) [[download](#)]
(<https://github.com/genkuroki/Statistics/raw/master/2022/11%20Hypothesis%20testing%20and%20confidence%20interval%20%20Two%20proportions.pdf>)]

定義された信頼区間

オッズ比の信頼区間

- Wald型の信頼区間
- Pearsonの χ^2 検定に付随する信頼区間
- Fisher検定に付随する信頼区間
 - Sterne型
 - Clopper-Pearson型

リスク比の信頼区間

- Wald型の信頼区間
- Pearsonの χ^2 検定に付随する信頼区間

比率の差の信頼区間

- Wald型の信頼区間

定義された確率分布

- Fisherの非心超幾何分布: $\text{FisherNoncentralHypergeometric}(m, n, r, \omega)$

関連講義動画

- [聴講コース 臨床研究者のための生物統計学](#) (<https://ocw.kyoto-u.ac.jp/course/328/>)
 - 佐藤俊哉, ランダム化ができないとき, 2018/10/25, 長さ 1:02:31) [<https://youtu.be/tUkyjZXU1vc>]
(<https://youtu.be/tUkyjZXU1vc>)]
 - 佐藤俊哉, 交絡とその調整, 2018/12/20, 長さ 1:00:52) [<https://youtu.be/ybdkQFEEdCPM>] (<https://youtu.be/ybdkQFEEdCPM>)]
 - 佐藤俊哉, 回帰モデルと傾向スコア, 2019/02/21, 長さ 1:04:44) [<https://youtu.be/cOHN444kBlo>]
(<https://youtu.be/cOHN444kBlo>)]

2.11.1 「P値と仮説検定と信頼区間(4) 母比率の違い」のノートブックの内容をJulia言語で直接実行するためのヒント

「検定と信頼区間(4) 比率の違い」のノートブック

(<https://nbviewer.org/github/genkuroki/Statistics/blob/master/2022/11%20Hypothesis%20testing%20and%20confidence%20interval%20%20Two%20proportions.ipynb>) の中の


```

a, b, c, d = 49, 965, 26, 854
@show a, b, c, d

$$\chi^2 = (a+b+c+d)*(a*d-b*c)^2/((a+b)*(c+d)*(a+c)*(b+d))$$

@show  $\chi^2$ 
pvalue = ccdf(Chisq(1),  $\chi^2$ )
 $\alpha = 0.05$ 
@show  $\alpha$ 
confint = find_zeros(0.5, 3.0) do  $\omega$ 
    A, B, C = 1- $\omega$ , a+d+ $\omega*(b+c)$ , a*d- $\omega*b*c$ 
     $\delta = 2C/(B + \sqrt{B^2 - 4A*C})$ 
     $\chi^2 = \delta^2 * (1/(a-\delta) + 1/(b+\delta) + 1/(c+\delta) + 1/(d-\delta))$ 
    ccdf(Chisq(1),  $\chi^2$ ) -  $\alpha$ 
end
@show confint;

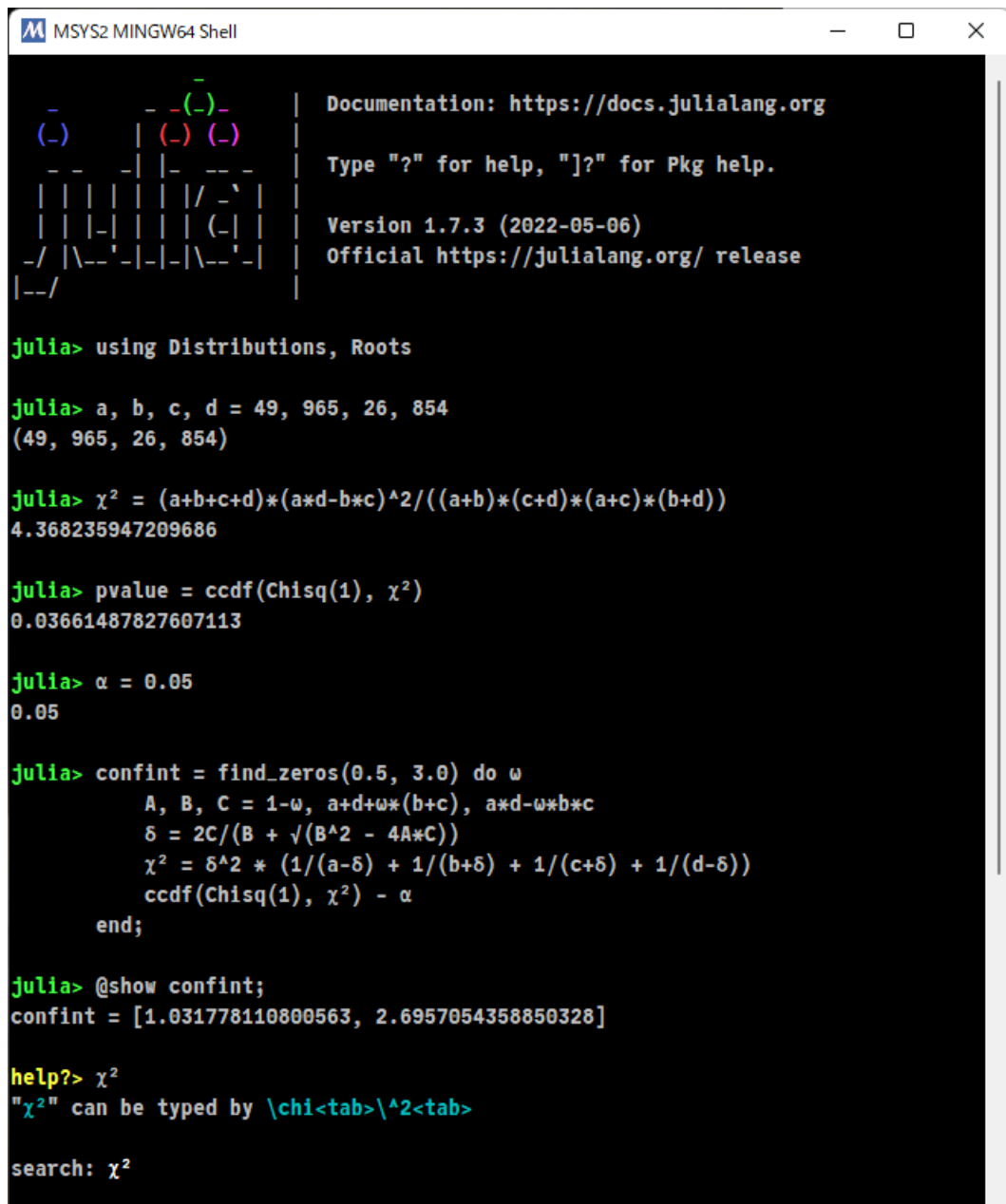
```

```

(a, b, c, d) = (49, 965, 26, 854)
 $\chi^2 = 4.368235947209686$ 
pvalue = ccdf(Chisq(1),  $\chi^2$ ) = 0.03661487827607113
 $\alpha = 0.05$ 
confint = [1.031778110800563, 2.6957054358850328]

```

の部分 directly Julia ([download \(https://julialang.org/downloads/\)](https://julialang.org/downloads/))で実行するためには以下のようにすればよい。



```

MSYS2 MINGW64 Shell

Documentation: https://docs.julialang.org
Type "?" for help, "]?" for Pkg help.
Version 1.7.3 (2022-05-06)
Official https://julialang.org/ release

julia> using Distributions, Roots

julia> a, b, c, d = 49, 965, 26, 854
(49, 965, 26, 854)

julia>  $\chi^2 = (a+b+c+d)*(a*d-b*c)^2/((a+b)*(c+d)*(a+c)*(b+d))$ 
4.368235947209686

julia> pvalue = ccdf(Chisq(1),  $\chi^2$ )
0.03661487827607113

julia>  $\alpha = 0.05$ 
0.05

julia> confint = find_zeros(0.5, 3.0) do  $\omega$ 
    A, B, C = 1- $\omega$ , a+d+ $\omega*(b+c)$ , a*d- $\omega*b*c$ 
     $\delta = 2C/(B + \sqrt{B^2 - 4A*C})$ 
     $\chi^2 = \delta^2 * (1/(a-\delta) + 1/(b+\delta) + 1/(c+\delta) + 1/(d-\delta))$ 
    ccdf(Chisq(1),  $\chi^2$ ) -  $\alpha$ 
end;

julia> @show confint;
confint = [1.031778110800563, 2.6957054358850328]

help?>  $\chi^2$ 
" $\chi^2$ " can be typed by \chi<tab>\^2<tab>

search:  $\chi^2$ 

```

注意すべきポイントは using Distributions, Roots を忘れないことである。

しかし、どのくらい前もって using しておけばよいのか分かり難いので、上のノートブックの上の方にある using 達をすべて前持って実行しておくという手もある。using には少し時間が取られるので、そのようにしたJuliaの窓をそのまま閉じずに放置しておき、超高級電卓として利用し続けると便利だと思われる。

χ^2 などの入力の仕方は、プロンプト julia> で ? と入力して、ヘルプモードのプロンプト help?> を表示させて、 χ^2 を貼り付けてエンターキーを押せばよい。その結果

```
"χ²" can be typed by \chi<tab>\^2<tab>
```

と表示されている。すなわち、\chi と入力してタブキーを押し、さらに \^2 と入力してタブキーを押せば χ^2 と入力できる。

2.12 12 P値と仮説検定と信頼区間(5) 母平均の差

- [Jupyter notebook](#) (<https://github.com/genkuroki/Statistics/blob/master/2022/12%20Hypothesis%20testing%20and%20confidence%20interval%20%20Two%20means.ipynb>) [[nbviewer](https://nbviewer.org/github/genkuroki/Statistics/blob/master/2022/12%20Hypothesis%20testing%20and%20confidence%20interval%20%20Two%20means.ipynb) (<https://nbviewer.org/github/genkuroki/Statistics/blob/master/2022/12%20Hypothesis%20testing%20and%20confidence%20interval%20%20Two%20means.ipynb>)]
- [pdf](#) (<https://github.com/genkuroki/Statistics/blob/master/2022/12%20Hypothesis%20testing%20and%20confidence%20interval%20%20Two%20means.pdf>) [[download](#) (<https://github.com/genkuroki/Statistics/raw/master/2022/12%20Hypothesis%20testing%20and%20confidence%20interval%20%20Two%20means.pdf>)]

定義された信頼区間

2群の平均の差の信頼区間

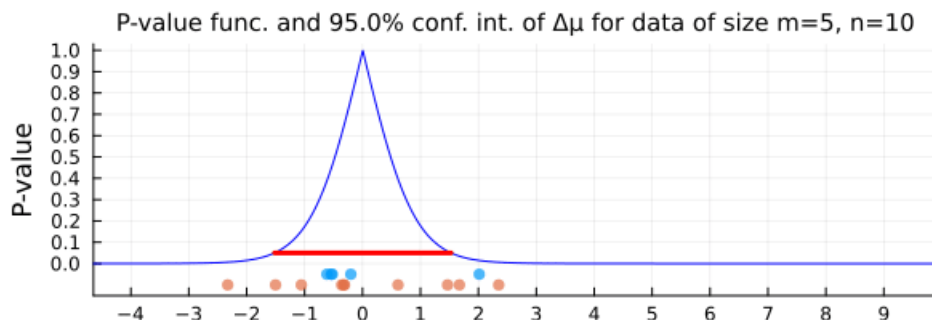
- Welchのt検定に付随する信頼区間

Studentのt検定は使用しない。

作成された動画

[データサイズを大きくしたときのP値関数と信頼区間の挙動](#)

(https://github.com/genkuroki/Statistics/blob/master/2022/images/confint_of_diffmeans.gif)



2.13 13 誤用を避けるための注意

- [Jupyter notebook](#) (<https://github.com/genkuroki/Statistics/blob/master/2022/13%20How%20to%20avoid%20misuse.ipynb>) [[nbviewer](https://nbviewer.org/github/genkuroki/Statistics/blob/master/2022/13%20How%20to%20avoid%20misuse.ipynb) (<https://nbviewer.org/github/genkuroki/Statistics/blob/master/2022/13%20How%20to%20avoid%20misuse.ipynb>)]
- [pdf](#) ([download](#) ())

P値に関する誤用を避けるために最初に読むべき基本文献は

- [統計的有意性とP値に関するASA声明](#) (<https://www.biometrics.gr.jp/news/all/ASA.pdf>)

である。ただし、以下の部分の「データ」は「統計モデル内でランダムに生成された仮想的なデータ」の意味であると解釈する必要がある。(「現実において観察されたデータ」であると解釈してはいけない。)

2. P 値とは？

おおざっぱにいうと、P 値とは特定の統計モデルのもとで、データの統計的要約（たとえば、2 グループ比較での標本平均の差）が観察された値と等しいか、それよりも極端な値をとる確率である。

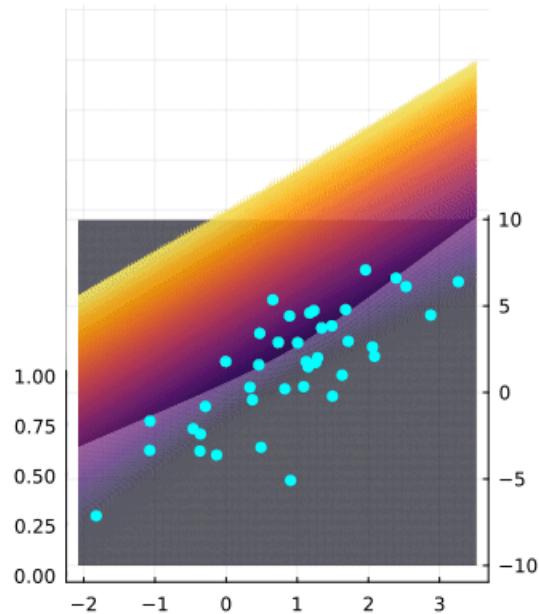
一連のノート群では、通常の統計学入門の教科書とは違って、「統計モデル」という用語を多用した。このP値の定義の具体例については他のノートで詳しく説明した。

2.14 14 回帰

- Jupyter notebook (<https://github.com/genkuroki/Statistics/blob/master/2022/14%20Regression.ipynb>) [[nbviewer](https://nbviewer.org/github/genkuroki/Statistics/blob/master/2022/14%20Regression.ipynb) (<https://nbviewer.org/github/genkuroki/Statistics/blob/master/2022/14%20Regression.ipynb>)]
- pdf (<https://github.com/genkuroki/Statistics/blob/master/2022/14%20Regression.pdf>) [[download](https://github.com/genkuroki/Statistics/raw/master/2022/14%20Regression.pdf) (<https://github.com/genkuroki/Statistics/raw/master/2022/14%20Regression.pdf>)]

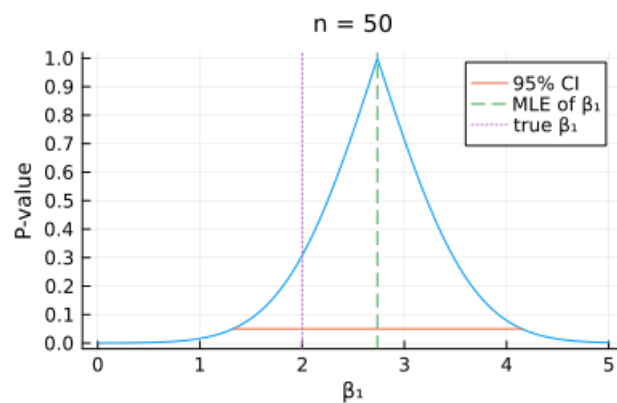
線形回帰の信頼区間に対応するP値函数

(https://github.com/genkuroki/Statistics/blob/master/2022/images/anim_linreg_confint_pvalfunc_3d.gif) (x軸は-2から2まで)



ロジスティック回帰の β_1 の信頼区間に対応するP値函数

(https://github.com/genkuroki/Statistics/blob/master/2022/images/anim_logisticreg_beta1_confint_pvalfunc.gif)



2.15 15 まとめ

- Jupyter notebook (<https://github.com/genkuroki/Statistics/blob/master/2022/15%20Summary.ipynb>) [[nbviewer](https://nbviewer.org/github/genkuroki/Statistics/blob/master/2022/15%20Summary.ipynb) (<https://nbviewer.org/github/genkuroki/Statistics/blob/master/2022/15%20Summary.ipynb>)]
- pdf (<https://github.com/genkuroki/Statistics/blob/master/2022/15%20Summary.pdf>) [[download](https://github.com/genkuroki/Statistics/raw/master/2022/15%20Summary.pdf) (<https://github.com/genkuroki/Statistics/raw/master/2022/15%20Summary.pdf>)]

1. 二項分布モデルでのClopper-Pearsonの信頼区間
2. P値函数について
3. Welchの t 検定について
4. 数学的な補足: 大数の法則と中心極限定理について

3 参考資料

3.1 以上で解説できなかったことに関するノート

- Kullback-Leibler情報量とSanovの定理 (2016-2018) [[pdf \(https://genkuroki.github.io/documents/20160616KullbackLeibler.pdf\)](https://genkuroki.github.io/documents/20160616KullbackLeibler.pdf)]
- ガンマ分布の中心極限定理とStirlingの公式 (2016-2018) [[pdf \(https://genkuroki.github.io/documents/20160501StirlingFormula.pdf\)](https://genkuroki.github.io/documents/20160501StirlingFormula.pdf)]
- 一般化されたLaplaceの方法 (2016) [[pdf \(https://genkuroki.github.io/documents/20161014GeneralizedLaplace.pdf\)](https://genkuroki.github.io/documents/20161014GeneralizedLaplace.pdf)]
- 最尤法とカイ二乗検定の基礎 (2017-2020) [[pdf \(https://genkuroki.github.io/documents/IntroMLE.pdf\)](https://genkuroki.github.io/documents/IntroMLE.pdf)]
- 確率論入門 (2017) [[pdf \(https://genkuroki.github.io/documents/IntroProbability.pdf\)](https://genkuroki.github.io/documents/IntroProbability.pdf)]
- ベイズ統計入門 (手書きのノート, 2019) [[pdf \(https://genkuroki.github.io/documents/2019-09-03_BayesianStatistics.pdf\)](https://genkuroki.github.io/documents/2019-09-03_BayesianStatistics.pdf)]
- ベイズ統計の枠組みと解釈について (2019) [[Jupyter notebook \(https://nbviewer.org/github/genkuroki/Statistics/blob/master/Introduction%20to%20Bayesian%20Statistics.ipynb\)](https://nbviewer.org/github/genkuroki/Statistics/blob/master/Introduction%20to%20Bayesian%20Statistics.ipynb)]
- Kullback-Leibler情報量と記述統計 (2016-2020) [[Jupyter notebook](#)]