

毒性学のためのR分析

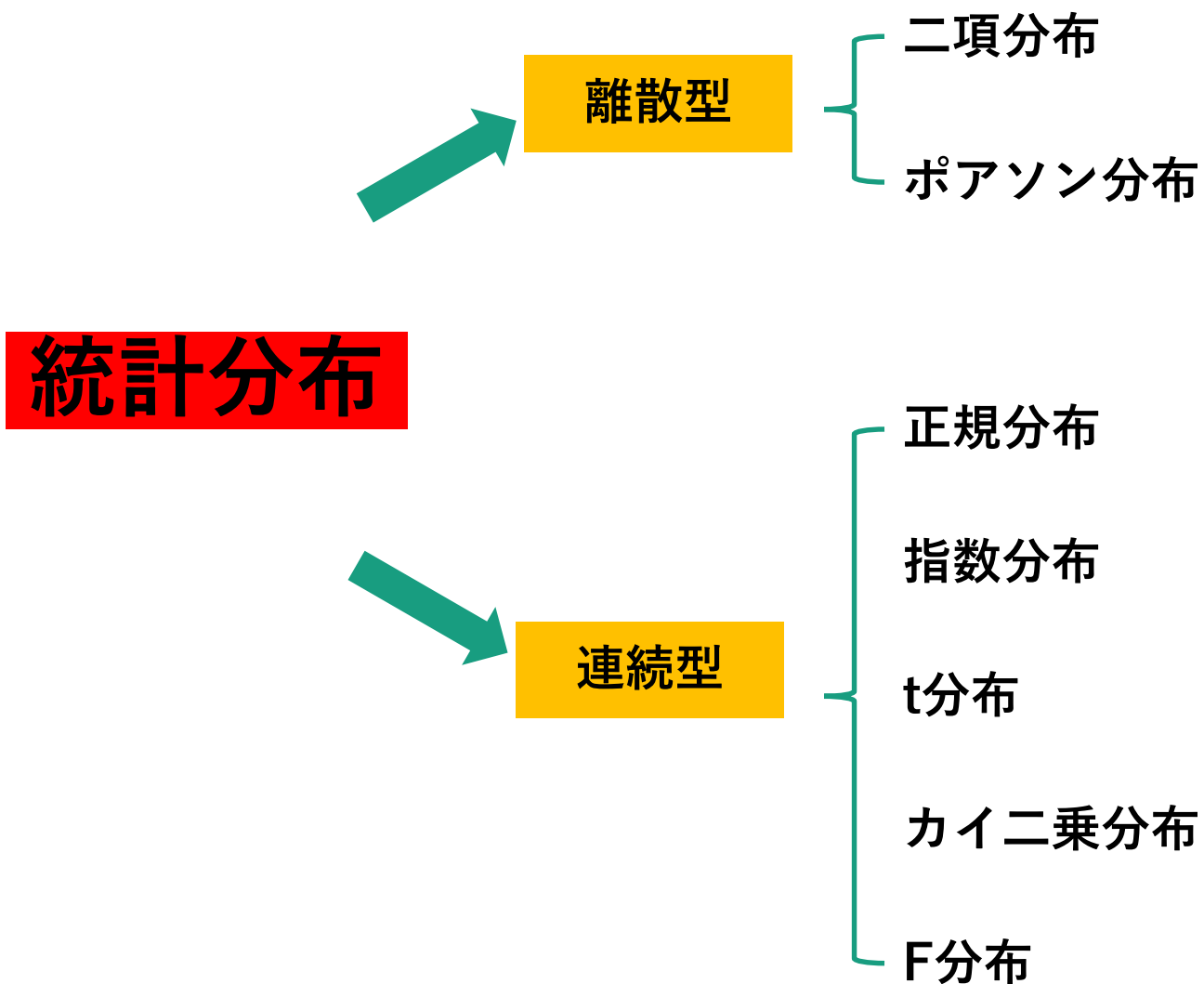
4. 基礎統計学（1）分布



主題

■ データ分析に必要な様々な統計分布を理解する

基本的な分布の種類



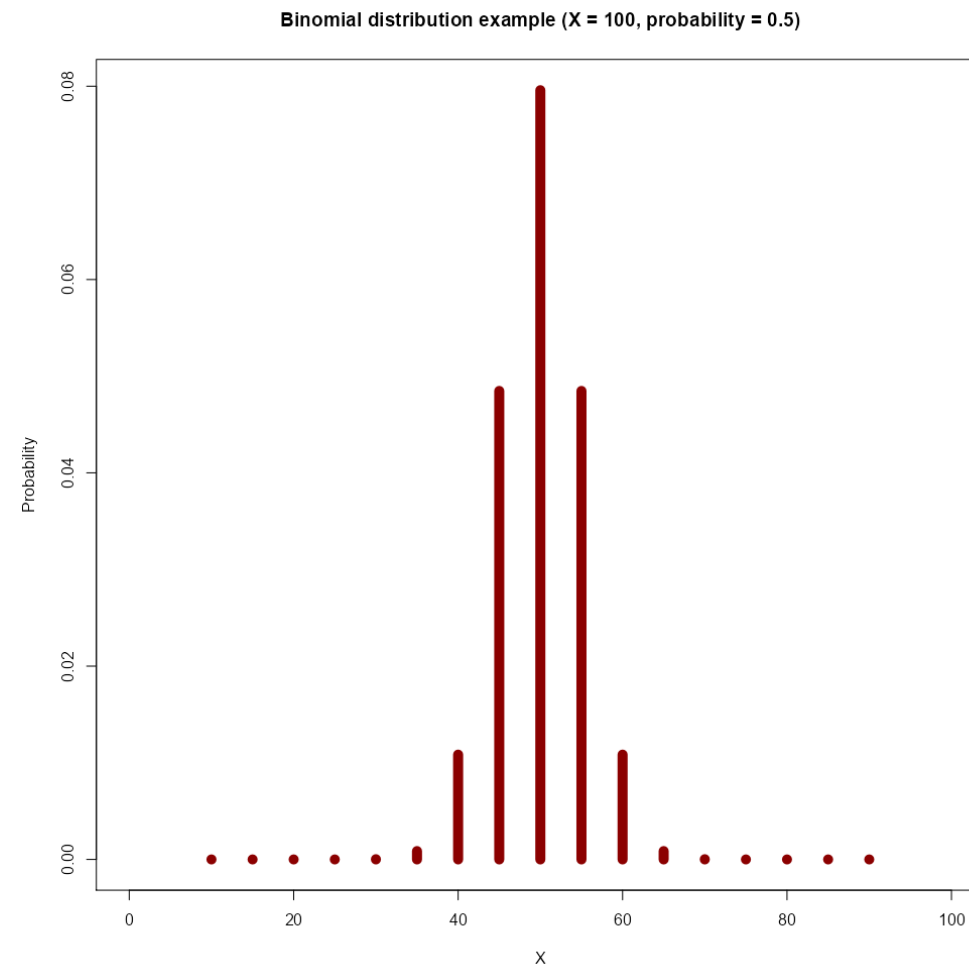
二項分布

■ 成功率が p のベルヌーイ試行を n 回



二項分布

例) 二項分布分布のグラフ

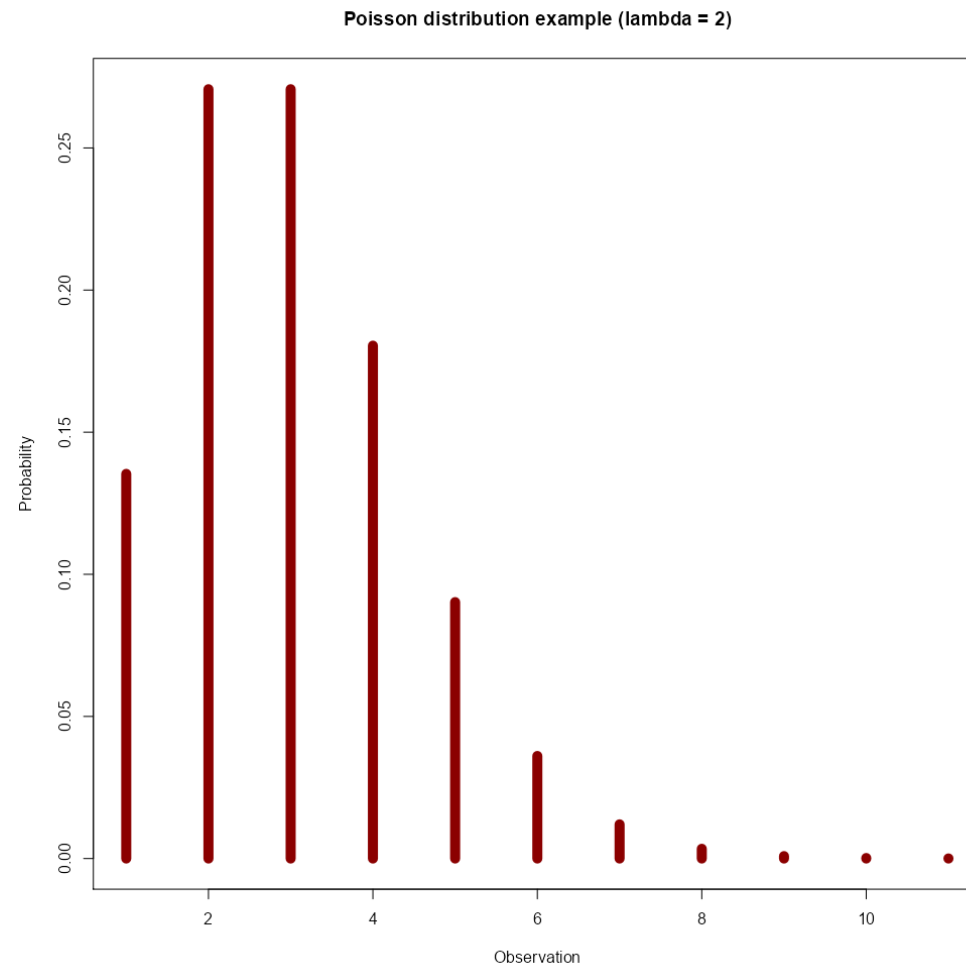


ポアソン分布

- 二項分布の計算のために開発した指数式を使った関数
- X が二項分布 $B(n, p)$ で、 $np = \lambda$ にし、 n が十分大きくて、 p が0に近いとき、ポアソン分布に近くなる
- 一定の時間、空間で発生するとある事件がランダムに発生するとき使う

ポアソン分布

例) ポアソン分布のグラフ



ポアソン分布

■ 例) とある店の支店の一時間当たりのお客さんの数が $\lambda = 30$ のポアソン分布に従う。1時間当たりのお客さんの数が27人の確率は？

```
> dpois(x = 27, lambda = 30)*100  
[1] 6.553248
```

約6.55%

正規分布

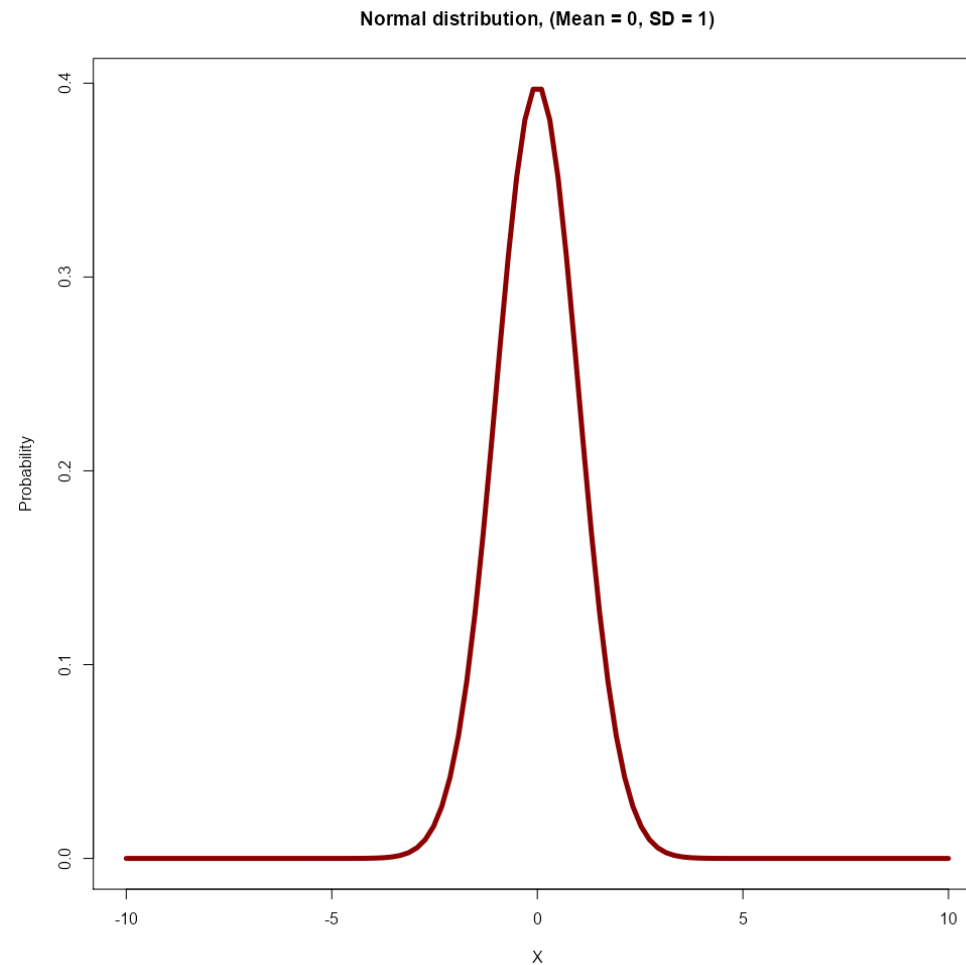
- 統計的推定、分析、検定の基本となる分布

- サンプルのサイズが大きくなればなるほど他の分布も正規分布に近くなる（中心極限定理）（普通30以上の場合）

- 平均 0、標準偏差 1 の正規分布は標準正規分布と呼ぶ

正規分布

■ 例) 正規分布のグラフ



指数分布

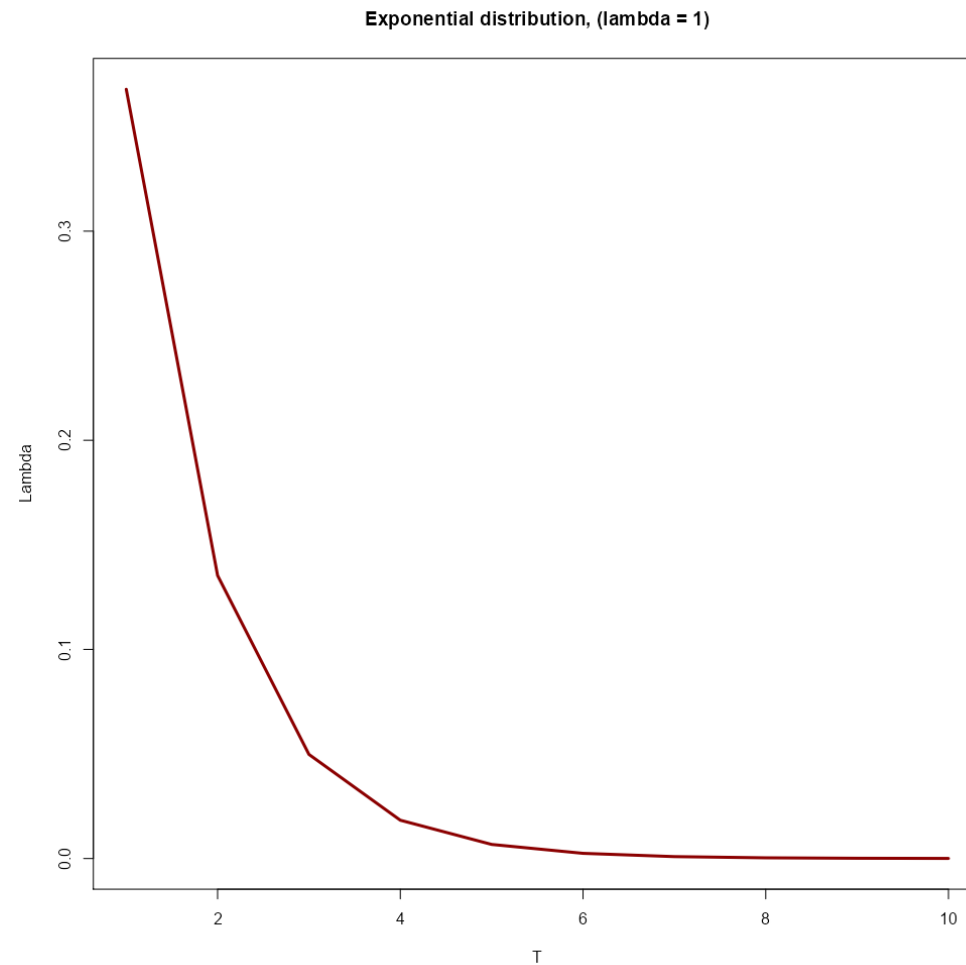
- とある事件が発生するまでかかる時間

- ポアソン分布と間違わないように要注意！

- 例えばA企業の洗濯機が壊れるまでかかる時間

指数分布

例) 指数分布のグラフ



T-分布

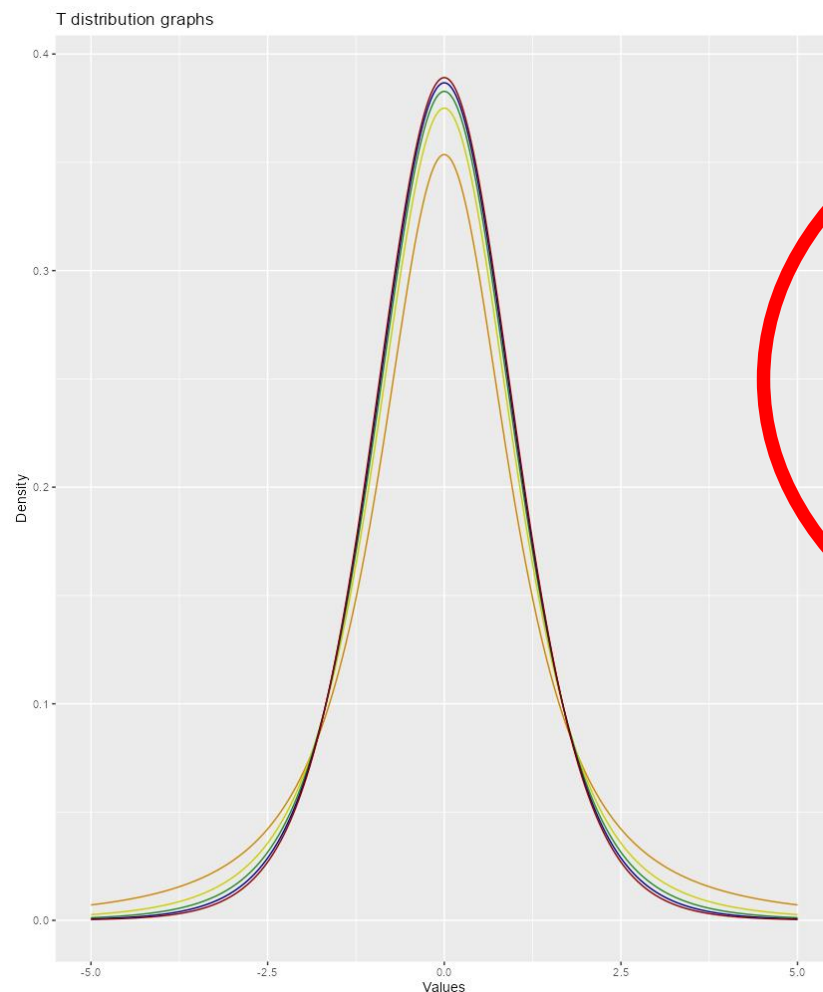
- 母集団の分散を知らないか標本の数が少ないとき使う

- 母平均の推定に普通に使われる

- 標本数が大きくなればなるほど正規分布に近くなる ($N > 30$)

T-分布

例) T-分布と正規分布のグラフの比較



colour

- T distribution df = 10
- T distribution df = 2
- T distribution df = 4
- T distribution df = 6
- T distribution df = 8

引数として自由度を使う

F-分布

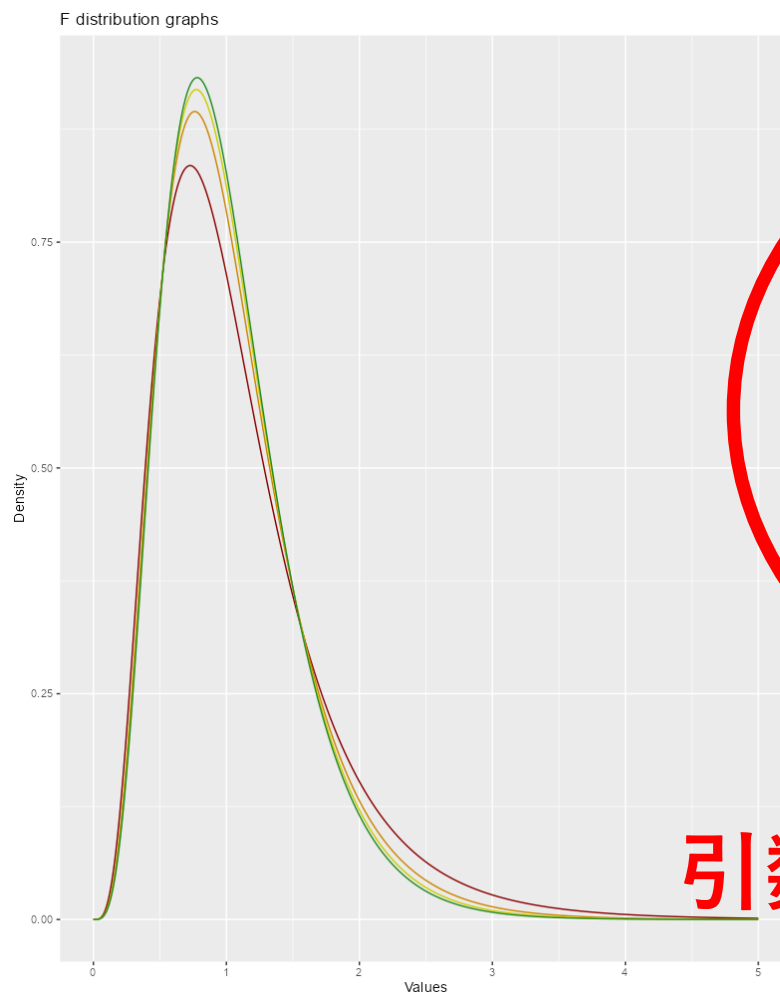
- 分散の類似性の推定に普通に使われる

- ANOVA分析に使われる

- 正数のみ + 非対称

F-分布

例) F-分布グラフの比較



colour

- F distribution df = 10 and 20
- F distribution df = 10 and 40
- F distribution df = 10 and 60
- F distribution df = 10 and 80

引数として二つの自由度を使う

カイ二乗分布

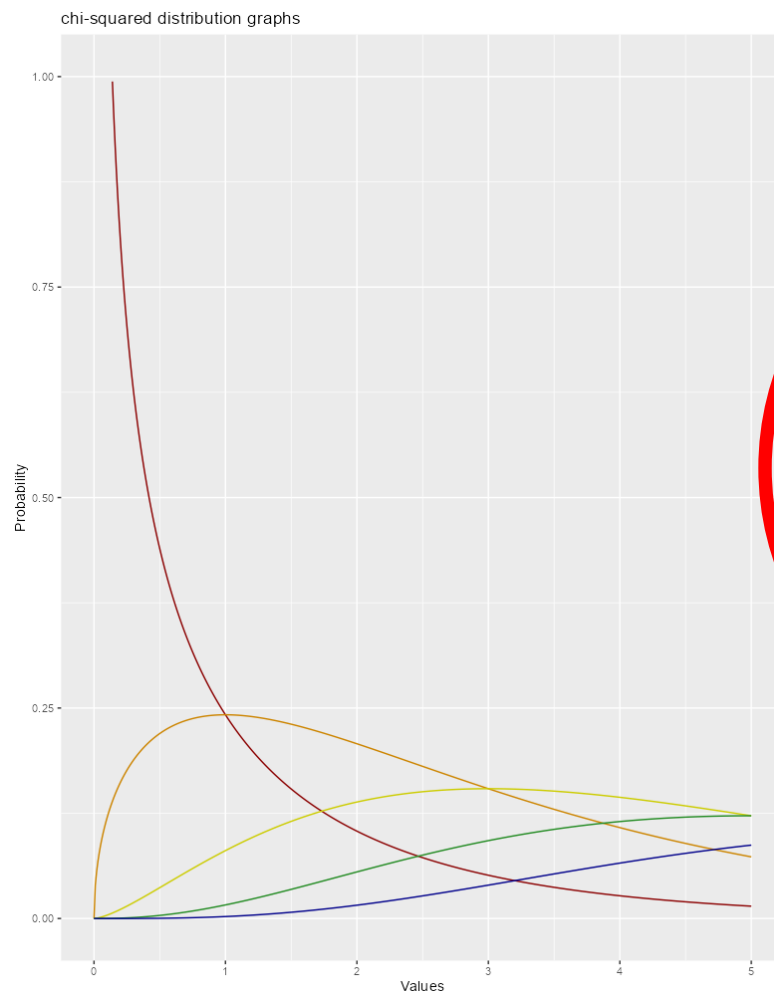
- 母集団の分散の推定、検定に使う

- カテゴリー型のデータの適合性、独立性に普通使う
例) 男性と女性が好きなデザートの種類の違い

- 正数のみ + 右寄り

カイ二乗分布

例) カイ二乗分布グラフの比較



colour

- chi-squared distribution df = 1
- chi-squared distribution df = 3
- chi-squared distribution df = 5
- chi-squared distribution df = 7
- chi-squared distribution df = 9

引数として自由度を使う

参考文献

- <https://rfriend.tistory.com/> (韓国語)
- <https://rvisuall.tistory.com/133> (韓国語)
- <https://www.geeksforgeeks.org/exponential-distribution-in-r-programming-dexp-pexp-qexp-and-rexp-functions/>