



χ^2 検定

勉強会 2023/06/16

M2 戸田梨鈴

今日の 内容

01

パラメトリックとノンパラメトリック

02

χ^2 検定

03

1変量の χ^2 検定（適合度の検定）



04

2変量の χ^2 検定（独立性の検定）

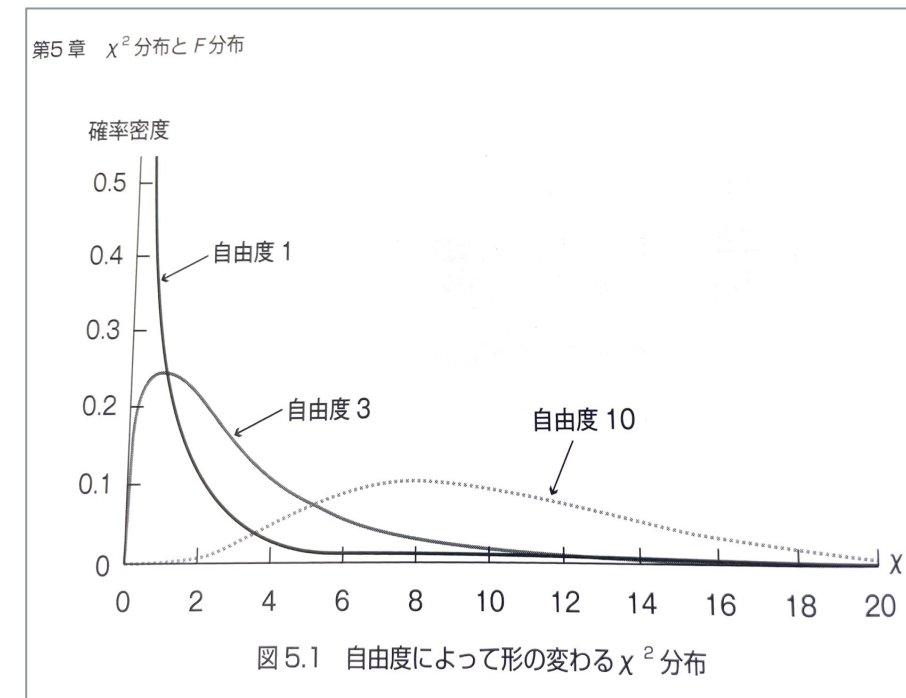
05

マクネマー検定（対応のあるデータ）

パラメトリックとノンパラメトリック

- パラメトリック： **parameter (母数) に基づくデータ**を想定して行う分析
= 特定の分布 (≒ 正規分布) を仮定する分析
 - **比率尺度**：ゼロを基準に、間隔や比率に意味があるもの (e.g., 身長)
 - **間隔尺度**：目盛りが等間隔で、間隔に意味があるもの (e.g., 気温) 量的データ
- ノンパラメトリック：特定の分布を仮定しない
 - **名義尺度**：他と区別し分類するためのもの (e.g., 性別)
 - **順序尺度**：順序や大小には意味があるが、間隔には意味がないもの (e.g., 順位) 質的データ
- サンプル数が小さい … 正しく分布曲線を描けない
- 外れ値がある … 分布曲線が歪む

- χ^2 検定(chi-square test) : χ^2 分布を利用する分析
 - χ^2 分布 : 各データの2乗の和 ($=\chi^2$ 値) が従う分布
 - ➡ 自由度によってグラフの形が変わる
 - ➡ 自由度が大きくなると正規分布に近づく



- 【補足】 自由度df : 代表値や合計値がある時に自由に決められる数
 - 実際のデータを表すものではなく **データの特徴**を示す
 - $m \times n$ のクロス集計表では $(m-1) \times (n-1)$

	○	×	合計
実験群	15	5	20
統制群	10	10	20
合計	25	15	40

1変量の χ^2 検定（適合度の検定）

- 1変量の χ^2 検定（適合度の検定）：得られたデータが**特定の分布（理論値）に適合しているか**
 - ある質問に対して、5名が「はい」15名が「いいえ」と答えた
この時、「はい」より「いいえ」の人の方が有意に多いと言えるか？
➔ **理論値**（はい50%, いいえ50%）に適合していない？（ $p < .05$ = 理論値とは適合していない）
- 使う関数
 - `factor()`：順序なし因子（名義尺度の形式）に変換
 - `str()`：データフレームの中身を表示（← RStudio 右上でも確認できる）
 - `labels=c('ラベル名1', 'ラベル名2', ...)`：ラベルをつける（**元のデータが小さい順に**）
 - `table()`：要素ごとの数を数えて、カテゴリ別に表にする
 - `chisq.test()`： χ^2 検定
 - `pie()`：円グラフの作成

1変量の χ^2 検定（適合度の検定）

- 理論値を指定して適合度の検定を行う場合

- 日本人の血液型の分布はA型が40%、O型が30%、B型が20%、AB型が10%であるとされる。今回得られた100人のデータは、この日本人の血液型分布と同じと言えるか？

血液型	A型	O型	B型	AB型	計
実測値	55	22	16	7	100
理論値	40	30	20	10	100

- `chisq.test()`で期待値を指定

`p=c()`で直接入力することができる

※ 合計が1になる割合の形で入力する

```
> x <- c(55,22,16,7)
> chisq.test(x, p=c(0.4,0.3,0.2,0.1))

Chi-squared test for given probabilities

data:  x
X-squared = 9.4583, df = 3, p-value = 0.02378
```

2変量の χ^2 検定（独立性の検定）

- 2変量の χ^2 検定（独立性の検定）：各要因は**互いに独立している（関係がない）**か
 - ある質問を男子20名、女子20名に対して行った。この質問に対する回答の性別による違いは有意であると言えるか？
 - ➡ 2つの要因（回答&性別）は独立でない？（ $p < 0.05$ = 独立でない）
- 使う関数
 - `table()`：`data$見出し` で使用する列を指定できる
 - `chisq.test()`： χ^2 検定
 - `matrix()`：行列を作る、`nrow`= 行数（縦）、`ncol`=列の数（横）
 - `rownames()`：行ラベル（上から順に入力）
 - `colnames()`：列ラベル（左から順に入力）

2変量の χ^2 検定（独立性の検定）

- 独立性の検定の注意点： χ^2 値を χ^2 分布に当てはめて得られる**近似した確率**
 - ➡ データ数が少ないと正しい値が求められない
 - ➡ **フィッシャーの直接確率検定** (Fisher's exact test)
 - 度数が0のセルがある
 - 度数が5以下のセルが、セル全体の20%以上
 - 周辺度数に10以下のものがある
- 使う関数
 - `fisher.test()`：フィッシャーの直接確率検定

マクネマー検定（対応のあるデータ）

- マクネマー検定：対応のある2値データについて、2つの結果に差があるか
 - ➡ 被験者内要因、介入前後の比較など
 - 子どもに「ゆるす」「ゆるさない」の2種類のストーリーを読み聞かせた。各ストーリーにおいて、主人公は違反者と「一緒に遊ぶ」「別々に遊ぶ」のどちらだと思うか尋ねた。ストーリーによって、児の選択が変わったと言えるか？
- 使う関数
 - `mcnemar.test()`：マクネマー検定
 - `correct=` ：イェーツの補正、サンプル数が小さい時は補正を行う（=T）

- ある幼稚園で、子ども200名（男児 100名、女児 100名）に3種類のおもちゃ（A, B, C）の中からどれで一番遊びたいかを選択させた。
 - この質問に対する回答と 児の性別に有意な関連は見られるか？
 - ① データの読み込み （データ： chap4practice.csv）
 - ② データ型の確認
 - ③ 表の作成（※ csvファイルには ID, gender, choiceについてのデータが含まれています）
 - ④ 検定手法の選択

※ 実際には この後に多重比較が必要

