

統計学(基礎)

第1回

看護・保健データと統計学のもろもろ

講師

- 高柳良太
 - 川崎市立看護大学看護学部 准教授
 - 東京学芸大学 教育学部 学校教育選修卒業
 - 東京学芸大学大学院 教育学研究科修士課程 学校教育専攻修了
 - 小学校教諭専修免許 幼稚園教諭2種免許 保育士
 - 司書 司書教諭
 - 知的財産管理技能士2級(管理業務)

資料について

- Teamsの「ファイル」の「クラスの資料」フォルダ内に、回数毎のフォルダを作成して置いておきます。
- 講義スライドは、配信で使用しているのと同じレイアウトのものと、1ページ4枚スライドの2タイプがあります
 - 必要に応じて使用してください
- データ等があるときも、そこに置きます

講義変更の予定

- 10/11 3、4限連続配信(14:10-15:50 16:00-17:40)
 - ※11/1が休講(岡田先生の講義が連続)
- 10/25 オンデマンド(収録)配信
 - 大学院オープンキャンパス実施日
- 11/1 休講
- 11/15 オンデマンド(収録)配信
 - 学部推薦入試
- 11/22 オンデマンド(収録)配信
 - 所用
 - 年来は12/20まで、年明けの1/10が14回目で最終講になる予定

シラバス

講義回	日付	内容
1回	9月27日	統計学の基礎事項の確認とJASPの動作確認
2回	10月4日	データセット設計
3回	10月11日	基本統計量とグラフ出力、JASPの出力利用
4回	10月11日	クロス集計表と χ^2 乗検定、フィッシャーの正確確率検定
5回	10月18日	対応のあるデータとマクネマーの検定
6回	10月25日	量的変数の順序化と度数分布表・ヒストグラム
7回	11月8日	母集団の等分散性の判断と2群の平均値の差の検定
8回	11月15日	対応のあるt検定
9回	11月22日	一元配置分散分析と多重比較
10回	11月29日	ノンパラメトリック検定
11回	12月6日	相関と相関係数の検定
12回	12月13日	重回帰分析
13回	12月20日	多変量解析法概説(予定)
14回	1月10日	ベイズ統計概説(予定)

講義の目的

- 統計解析手法の意味を知る
 - 論文で出てきた手法の意味がなんとなくわかるようになる
 - 分析に応じて必要なデータの種類のわかるようになる
- 自分で基本的な解析ができるようになる
 - 必要なデータを集められる
 - 分析アプリを実行することができる
 - アプリが出した結果を理解することができる

目標

- 何がわからないのかが、わかるようになる
 - 手法の意味がわからない
 - 結果の読み方がわからない
 - アプリの使い方がわからない

この授業について

- 計算は原則しない
 - アプリに任せる
- 数式等は理解してくれると嬉しいがそこまで求めない
 - それよりは、手法の違いと結果の読み方をわかることが先決
- 評価
 - 自分でデータが集められて(設計ができて)、分析手法が選択できて、分析結果の解釈ができる

軽く復習

そもそも統計とは

統計とは

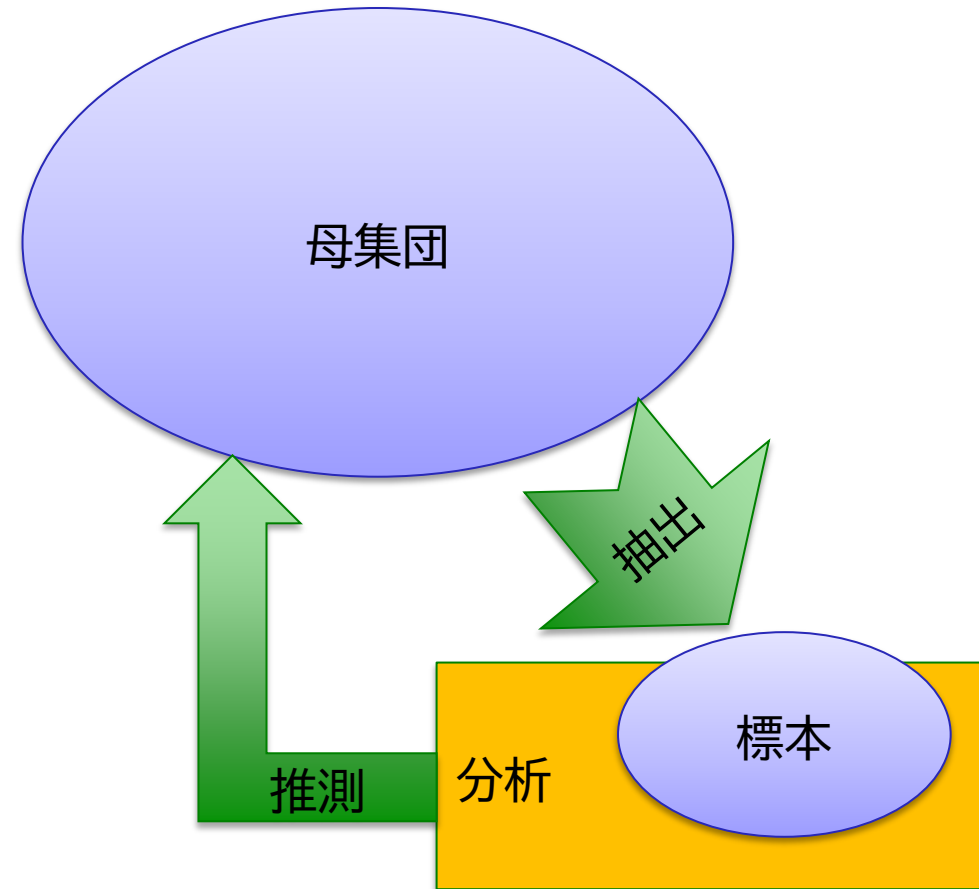
- 数字で表されている情報について、計算や表、グラフを使ってその意味を明らかにするための道具(ツール)
 - 計算結果や表、グラフのことを統計と呼び、計算(解析)そのものも統計とよぶことがある

2種類の統計

- 記述統計
 - 手元にあるデータについて、計算や表、グラフなどでその意味を説明する
- 推測統計
 - 手元にあるデータについて、計算や表、グラフなどを用いて、手元にはないデータを推定する

推測統計の基本的な考え

- 母集団と標本抽出
 - 全体(母集団)から偏りなく得られた(抽出された)データ(標本)を使えば、全体を測定しなくても全体をある程度の精度で推測できる
 - 手元にあるデータが抽出された標本であると仮定できれば、全体を推測できるとする
 - そもそも、対象となるデータは大きすぎて現実的にデータがとれない



なぜ統計が必要か

- 大量データを整理するため
 - 自分の理解や他人への説明、情報発信のため
- 得られたデータから全体を推測するため
 - すべてのデータを得ることが現実的でないため

統計解析で何ができるか

- ・ データの整理・要約をする
 - 統計量の算出
 - 基本統計量、表・グラフ作成
- ・ 比較する・違い(差)を知る
 - 検定
- ・ 関係や傾向を知る・予測をする
 - 相関・回帰
- ・ グループ分けをする
 - 多変量解析

ツールとして

統計解析アプリ

統計アプリ

- この授業では、統計解析はアプリを使用します
- メインでは無いけど、以下も説明します
 - アプリの使い方
 - 結果の見方
- 本来はどのアプリでもいいのですが、分析の指定方法や出力のレイアウトが異なるので、一応指定させてください

主な統計アプリ

- SAS
- SPSS
- S(S-PLUS)
- R
 - EZR
 - JASP
 - JAMOVI
- Python

R系

- EZR
- JASP
- JAMOVI
- R言語にGUI画面をかぶせている
 - バックでは同じRが動いている
 - どれもフリーソフト(R自体が無償だから)

本講義で使用するアプリ

- JASPがお勧め
 - Windows Mac Linuxで動く
 - データ管理の見た目がExcelっぽい
 - SPSSに見た感じ似てる
 - データライブラリが充実していて勉強になる
 - 「JASP ダウンロード」で検索すると出てくる

データのフォーマット

- Excelで準備すれば大丈夫
 - CSV形式にすればベストだけど、Excelでも平気かな
 - 日本語はない方がいい
- JASPのデータ編集機能はあまり使わない
 - 機能がしょぼい

データのフォーマット(形式)

- Excelで準備する場合
 - 縦(列)が変数
 - 横(行)がケース
 - 1セルに一つのデータ

JASPについて

JASPとは

- JASP(Jeffreys's Amazing Statistics Program)
- <https://jasp-stats.org/>
- アムステルダム大学心理学部で開発されているオープンソースプロジェクト
 - Free
 - Friendly
 - Flexible




JASPの特徴

- 頻度論とベイズ統計ができる
 - 頻度論的統計: 古典的な統計解析
相対度数(頻度論的確率)にもとづき、p値を用いる統計手法。
 - ベイズ統計: 個人的信念(主観確率)にもとづき、パラメータの事後分布やベイズファクターを用いる統計手法。
 - ※ Harold Jeffreys ベイズ統計で用いられる事前分布の一つ
Jeffreys prior(ジェフエリー事前分布)の考案者
- この講義ではベイズ統計は原則扱わない

JASPを進める理由

- Windows Mac Linuxで動く
- データ管理の見た目がExcelっぽい
 - Excelデータの読み込みが比較的うまく行く
- SPSSに見た感じ似てる
 - SPSSデータの読み込みもなんとかなる
- データライブラリが充実していて勉強になる

JASPの変数型(列の種類)

- 名義 
 - カテゴリーデータともいう。度数を集計するだけで、四則演算できないデータ。英語表記だとNominal、オランダ語ではNominaalとなる。
- 順序 
 - 順番に意味がある名義データ。英語表記だとOrdinal、オランダ語では Ordinaal。
- スケール 
 - 量的データ。英語表記だとScale、オランダ語では Schaal。

データ記述 long型wide型

• long

No	回数	値
1	1	70
2	1	60
3	1	75
1	2	75
2	2	90
3	2	75
1	3	80
2	3	85
3	3	90

• wide

No	1回目	2回目	3回目
1	70	75	80
2	60	90	85
3	75	75	90

可読性はwide型の方がよいが、統計解析アプリでは、long型でデータを持っている方が扱いやすい。

JASPのデータライブラリ

- クリックの仕方で、内容が変わる

The image displays two screenshots of the JASP Data Library interface, illustrating how different clicks lead to different content.

Top Screenshot: The interface shows the 'Data Library' (データ ライブラリ) section. A button labeled 'Category' (カテゴリ) has an arrow pointing to '1. Descriptives'. A callout box points to the 'Data Library' header, stating: 'ライブラリ名をクリックすると、データファイル(csv)が開く
データテーブルのみの表示になる' (Clicking the library name opens the data file (csv) and displays only the data table).

Bottom Screenshot: The interface shows the 'Data Library' (データ ライブラリ) section. A button labeled 'Category' (カテゴリ) has an arrow pointing to '1. Descriptives'. A callout box points to the 'JASP File' (JASPファイルを開く) button, stating: '説明部分をクリックすると、JASP形式のファイルが開く
説明や、サンプルの分析結果が表示される' (Clicking the description part opens the JASP format file and displays the explanation and sample analysis results).

JASPのデータライブラリ

データの著作権について

Most data sets contained in the JASP Data Library are already publicly available because they have been presented in academic articles and books. The legal consensus is that such data are facts and therefore do not fall under copyright protection. An exception concerns data that have not been collected but constructed in order to serve an educational purpose, as is the case for most data sets from Andy Field's popular course books.

(<https://jasp-stats.github.io/jasp-data-library/>)

データの著作権について

JASP データライブラリに含まれるほとんどのデータセットは、学術論文や書籍等で、すでに公開されたものです。法的なコンセンサスは、そのようなデータは事実であるため、著作権保護の対象にはなりません。

※高柳注:日本の著作権法では、第二条一において「著作物 思想又は感情を創作的に表現したものであつて、文芸、学術、美術又は音楽の範囲に属するものをいう。」としているので、単純なデータは著作物に当たらず著作権の保護対象にはならないという考えが一般的

データの著作権について

例外は、Andy Fieldの人気のあるコースブックのほとんどのデータセットの場合のように、収集されていないが、教育目的を果たすために構築されたデータに関するものです。

※高柳注:ライブラリの「Books」にあるデータは、Discovering Statistics Using IBM SPSS Statistics(Field,A.P.(2017))と Introduction to the Practice of Statistics(Moore,D.S.,McCabe,G.P.,& Craig,B.A(2012))という2冊の書籍からの転載となっています。こちらはCC BY-NC-ND(表示-非営利-改変禁止)となっており、非営利での使用は可能ですが、改変をしたものの再配布等は禁じられています。なおCC BY-NC-NDについては、クリエイティブコモンズの当該ページ(<https://creativecommons.org/licenses/by-nc-nd/4.0/deed.ja>)を参照のこと

おまけの話

- 公開されているデータには著作権がない
 - 「思想又は感情を創作的に表現したもの」ではないから
 - 同じ条件でやれば誰でも同じ結果になるはずだから
 - 著作権は創作物に対して発生するもので、労力に対して発生するものではないから
 - 書籍等で説明のために作ったダミーのデータは別(創作だから)
 - 保管されているデータファイル等を盗むとそれは窃盗罪

来週までのお願い

お知らせ

JASPのインストールと

- <https://jasp-stats.org/>
- 「JASP インストール」で検索しても大丈夫
- Windows、Macintosh、どっちでも可
- インストールしたら、ライブラリのデータ等で、ちょっといじってみておいってください。