

言語統計学 A  
Statistics in Linguistics A

---

2023 年度 春・夏学期

# 目次

第1講	はじめに：「仮説」と統計手法の選択.....	2
第2講	推測統計学の四つの重要概念.....	15
第3講	t検定（対応なし）.....	39
第4講	単回帰分析.....	69
第5講	重回帰分析.....	101
第6講	分散分析（対応なし）.....	143
第7講	線形混合効果モデル.....	201

## 学習の目標

- 研究とは何かを明確に述べることができる。
- リサーチクエスチョンとは一般にどのようなものかがわかり、自分の研究のリサーチクエスチョンを明確に述べることができる。
- リサーチクエスチョンには、探索的なものと検証的なものがあることがわかり、任意のリサーチクエスチョンがどちらに属するものなのか判断することができる。
- 仮説を明示的なものへと具現化させていく際に、自分が関係を見る対象たちを明示的に指摘でき、また、統計学ではそのような対象を変数（要因）と呼ぶことが分かる。
- 自分が抱いている仮説において、独立変数と従属変数を区別し、指摘することができる。
- 計算できる演算の種類に応じ変数を四つの尺度水準に分類でき、自分のデータがどの尺度に適しているか判断できる。
- データの測定に際し、リッカート尺度やマグニチュード推定法と呼ばれる収集手法があることがわかり、説明できる。
- 統計的研究には、探索的研究と検証的研究の二つのタイプのアプローチがあることがわかり、どのような状況でどちらを選択すべきか、説明できる。
- 独立変数と従属変数のタイプによって用いられる検証的手法が分類されるということが分かる。
- 統計モデルについて、「大変そうだけれどもやってみたい!」と思うことができる。

統計分析の具体的な基礎を細かく学んでいく前に、この第1講では、研究という視点と結びつけながら、今学期に学ぶ統計分析の概要をつかんでおきましょう。なんとなく受講するのではなく、自分の研究・リサーチクエスチョンを常に念頭において、それを解くための手段として統計的手法を位置づけ、最終的に何らかの研究成果を上げることを目標に、統計的手法のスキルを伸ばして行ってください。

(1) 研究

これは、過去の人類が作り上げた知識の体系を広げ、新しい知識の体系を作る試み。

(2) リサーチクエスト

これは、その研究で答えを明らかにする問いのこと。

① 探索的なリサーチクエスト (漠然／研究の始点)

明確な「仮説」を持たずに、現象の性質をデータから浮かび上がらせ、発見しようとする探索的な問いのこと。

※(現象があまりにも複雑すぎて) 先行研究などの知見からは「仮説」が立たないときに問われることが多い。



探索的なリサーチクエストの具体例

(問) どういう動詞が『と／の／こと』をどういうふうを選択しているのか

(答) A という動詞は〇〇という傾向がある。

B という動詞は△△という傾向がある。

⋮

動詞によって様々だ。ただし、□□という動詞に〇〇という傾向が強いようだ。

② 検証的なリサーチクエスト (焦点化された問)

先行研究などから見えてきた「仮説」の正しさを問うリサーチクエストのこと。



検証的なリサーチクエストの具体例

先行研究では、抽象的な命題では「こと」が、具体的な命題では「の」が選択されやすいという指摘がある (Kuno 1973)。

そこで、

(例1) 〇〇というコーパスを見ると、命題の抽象性で「の／こと」の選択が変わるのか、という傾向を研究する。

(例2) 命題の抽象性で「こと」を用いた文に対する、実験協力者の容認度が変わるのか、という傾向を研究する。

(1) 仮説 Hypothesis

① 「仮説」の定式化

仮説を明確にするための出発点は、「何と何の関係を明らかにしたいのか」を言語化すること。

② 変数 variable (要因 factor)

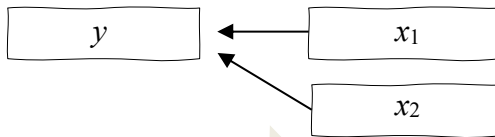
統計学では、仮説において関係を見る対象たちのことを「変数 (要因)」と呼ぶ。

※ 独立変数を「一つだけ」想定する場合：



「 $x$  が  $y$  に影響を与える」  
「 $x$  によって  $y$  は変わる」  
「 $x$  が増えれば (減れば)  $y$  はよく (悪く) なる」  
「 $x$  が増えれば (減れば)  $y$  は増える (減る)」

※ 独立変数を「複数」想定する場合：



「 $x_1$  と  $x_2$  が  $y$  に影響を与える」  
「 $x_1$  と  $x_2$  によって  $y$  は変わる」  
「 $x_1$  が増えれば (減れば / 一定で)  
 $x_2$  が増えれば (減れば / 一定なら)  
 $y$  はよく (悪く) なる」



変数の分類

- ① 独立 (予測) 変数 Independent (Predictor) variable  
これは、それを用いて何かを予測する変数のこと。  
SS
- ② 従属 (応答) 変数 Dependent (Outcome) variable  
これは、独立変数から値を予測される変数のこと。

## (2) 変数の尺度

変数はその数学的性質によって、次の四つに分類される。

### ① 名義尺度 Nominal scale

分類の区分を表す変数。

足し算も、大小の比較も行えない。

### ② 順序尺度 Ordinal scale

名義尺度に順序が付け加わった変数。

大小の比較は可能だが、足し算や掛け算は行えない。

### ③ 間隔尺度 Interval scale

順序尺度に足し算や引き算が付け加わった変数。

掛け算や割り算は行えない。

### ④ 比率尺度 Ratio scale

間隔尺度に掛け算や割り算が付け加わった変数。

四則演算が全てできる。

## (3) よくあるデータの収集方法

### ① リッカート尺度 Likert Scale

「5:とても自然 4:自然 ...」のような離散的な段階評価

不自然	1	2	3	4	5	自然
数字をクリックするか、数字キーを押してください。						

### ② マグニチュード推定法 Magnitude Estimation

Step 1

\_\_\_\_\_ 21  
\_\_\_\_\_ 7  
\_\_\_\_\_ 25

Step 2

\_\_\_\_\_ ー

Step 3

He am a good boy.

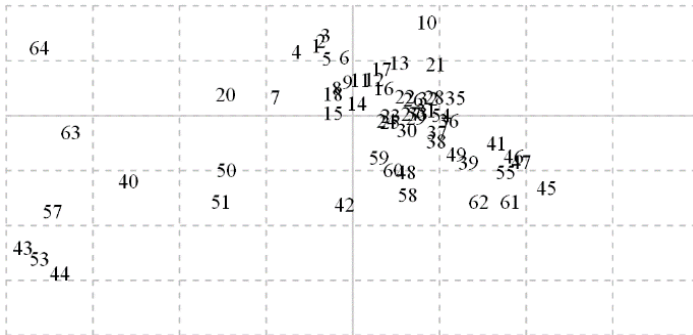
## 📖 ノート3 仮説と統計的アプローチの分類

### (1) 探索的統計手法：「仮説」の探索に有効

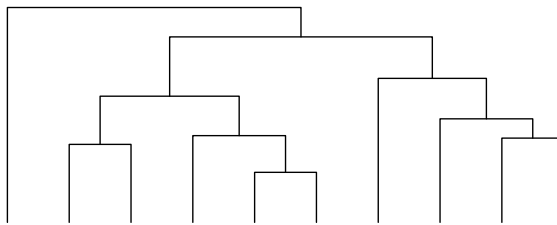
これは、どのような仮説が現象に対して適切かを発見するためにデータを整理する目的で利用される統計手法。

例：主成分分析、クラスター分析、多次元尺度構成法、…

#### ① 散布図を改良したもの



#### ② 樹形図に基づくもの



### 探索的統計手法の注意点

#### ① 運・不運がある

必ず何らかの価値のある発見ができるという保証はない。安定して論文を生産するうえでは、頼りすぎは危険。

#### ② 「標本」の整理がメインの場合が多い

母集団への推測がやりにくいものも多い。例えば、今回の樹形図／散布図が、母集団でも成立するかはわからない。

#### ③ 研究者の解釈への依存度が高い

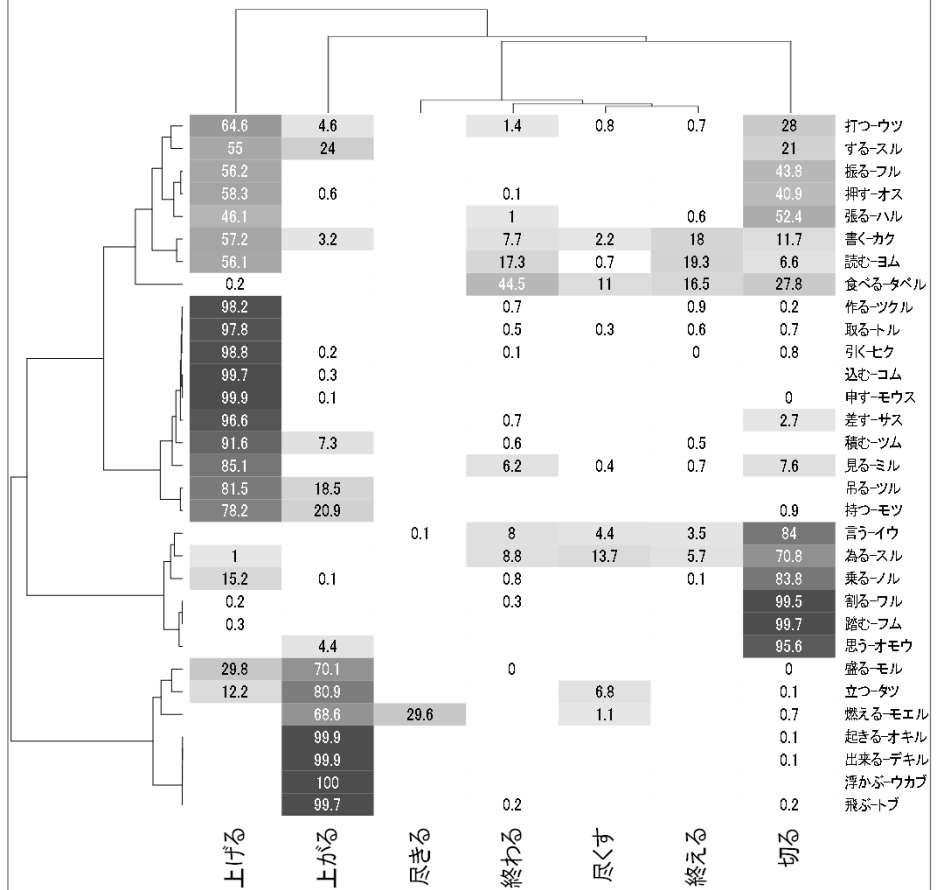
樹形図／散布図の解釈に伴って生じる研究者の価値判断への依存度が、検証的手法と比べると、大きい。



## コーパスを用いた探索的言語学研究の具体例

リサーチクエスチョン：複数存在する完了を表す複合動詞  
「動詞+V」にはどのような使い分けがあるのか？

	上げる	上がる	切る	尽くす	尽きる	終える	終わる
言う	0.000	0.000	0.840	0.044	0.001	0.035	0.080
浮かぶ	0.000	1.000	0.000	0.000	0.000	0.000	0.000
打つ	0.646	0.046	0.280	0.008	0.000	0.007	0.014
起きる	0.000	0.999	0.001	0.000	0.000	0.000	0.000
⋮							
読む	0.561	0.000	0.066	0.007	0.000	0.193	0.173
割る	0.002	0.000	0.995	0.000	0.000	0.000	0.003





(2) 検証的統計手法：「仮説」の検証に有効

これは、現象に対して立てられた仮説が適切かを検証するために利用される統計手法。

① 従属変数：比率尺度の変数 (☞言語統計学 A)

この場合、従属変数が正規分布に従うと仮定してモデル化することが多い。大きく次のような手法がある。

(A) 独立変数：固定効果のみ

独立変数の数 / タイプ	名義尺度	比率尺度
1つ	第3講	第4講
-----		
2つ以上	第6講	第5講

(B) 独立変数：固定効果 + 変量効果 (=混合効果)

独立変数の数 / タイプ	名義尺度 OR 比率尺度
1つ以上	第7講

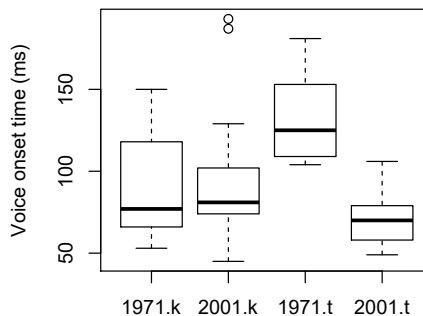


比率尺度を従属変数に持つ言語学研究の具体例

チェロキー語では、Voice onset time の長さに変化が生じているという指摘がある。「1971年と2001年 ( $x_1$ ) に、Voice onset time の長さ ( $y$ ) の平均に差は存在するのか」を、t, k など子音の違い ( $x_2$ ) も考慮して検証する

(Johnson 2008)。

⇒ 「音声産出実験」を計画



② 従属変数：比率尺度ではない変数 (☞言語統計学 B)

従属変数が 正規分布以外の分布 に従うと仮定

(A) 独立変数：固定効果のみ

---

独立変数の数 / タイプ 名義尺度 OR 比率尺度

---

1つ以上

(B) 独立変数：固定効果 + 変量効果 (=混合効果)

---

独立変数の数 / タイプ 名義尺度 OR 比率尺度

---

1つ以上



#### 名義尺度を従属変数に持つ言語学研究の具体例

日本語では、「走りません」「走らないです」のように否定丁寧形に二つの可能性がある。統計を使わない研究で終助詞が付くと、「ないです」形が選択されやすくなるかもしれないという指摘がある。そこで、「後ろに終助詞があるかないか (x) によって、どちらが選択されるのか (y) が影響を受けるのか」を、大規模なコーパスを用いて検証する。

⇒「コーパス観察研究」を計画



#### 名義尺度を従属変数に持つ文学研究の具体例

探索的研究の結果から、ある作家には、前期作品に色を用いた描写が多く見られるのに対し、後期ではほとんど色への言及はないという傾向が指摘されている。さらに自分は、「各文に色への描写が含まれているのか否か (y) は、作品の出版年 (x<sub>1</sub>) だけでなく、地の文か会話文か (x<sub>2</sub>) によって変わるのではないか」という仮説を立て、この仮説の是非を検証する。

⇒「コーパス観察研究」を計画