中間試験対策ノート

1. データとグラフ

全数調査と標本調査の違い

—全数調査をするには母集団全体を調べる必要があるため、莫大な時間と費用が掛かる。ただ、一番正確な結果を得られるので、他の統計の基本となる統計資料の作成など、正確さが費用より大事なときには全数調査を用いる。標本調査は母集団の一部分から標本を取って調査するので、費用が抑えられる。また、標本の抽出が無作為抽出によるものであれば、結果の正確度も母集団に近い。

データの種類

—データには横断面データ、時系列データ、パネルデータがあって、横断面データは調査の時点が固定されたデータであり、多数の標本をその調査項目ごとに表す。例としては国勢調査や民間給与実態統計調査などがある。時系列データは一つの指標を時間軸に沿って観測したデータであって、日経平均株価などのマクロ指標は主に時系列データである。パネルデータは時間の流れに沿って調査対象の変化を表すデータで、個人の社会・経済的な変化を追跡調査した慶応大学の日本家計パネル調査(JHPS)などがパネルデータに値する。

グラフの種類

—度数分布図、棒グラフ、円グラフ、散布図、折れ線グラフなどがある。各グラフの詳細は授業のスライドを参照

予想される問題

- A.「母集団」、「標本抽出」、「調査項目」などの統計調査の用語の理解度確認
- B.グラフの図を提示し、それがどのグラフであって、どの種類のデータから作成されたかを問う問題

2. 相関関係と因果関係

相関関係と相関係数

—相関関係とは、二つの変数の間の関係で、一方が変動するとき、他方もそれに伴って変動する関係性を言う。相関関係には正の相関と負の相関があって、一方がプラスに動くと他方もプラスに動くのが正の相関であり、一方がプラスに動くとき、他方はマイナスに動く逆関係が負の相関である。相関係数は相関関係を測る尺度で、-1から1までの間の値を取る。-1に近いほど、強い負の相関で、1に近いほど、強い正の相関、また、0に近いほど無相関を示す。

因果関係

—因果関係とは、「原因」と「結果」が分かれている相関関係で、原因が与えられたら、それに伴って結果が変化する方向性のある関係性である。原因だけを変えた別の平行世界を想像することで(反事実)、因果関係が成り立つかをざっくり検証できる。

因果関係の成立条件

—因果関係を確認するとき、3つの条件を検証する。一つ目は「全くの偶然ではないか」で、偶然によって見せかけの因果関係が成り立っていないかを検証する。二つ目

は「逆の因果関係は存在していないか」で、因果関係の原因と結果を入れ替えて成立するかを検証する。三つめは「第3の変数は存在していないか」で、原因と結果、両方に影響を与える交絡因子があるかを検証する。

ランダム化対照試験

一薬の臨床試験などに用いられる手法で、ランダムに原因となる要素を変動させて、その結果の違いから因果関係を導出する。「薬を飲む」という原因から「薬の効果」の結果への因果関係を想像すれば分かる。ランダムにする理由は、そうでないと標本抽出の無作為抽出の議論と同様に、間違った結果を導出しやすくなるからである。

自然実験

一偶然によって原因となる要素がランダムに変動させられた状況を自然実験と呼ぶ。費用と倫理的な問題で人に対して実験ができない社会科学においては、自然実験が因果関係の検証の代替策として好まれる。

予想される問題

- A. ある二つの変数を提示し、その間に因果関係が成り立つかを厳密に議論する。その場合、因果関係の成立条件となる3つの条件を用いる。
- B. ランダム化対象試験と自然実験の概念を問い、自然実験が社会科学において好まれる理由を記述する問題。

3. 自然実験の例

例を扱う章は、試験の対象としない。

4. 回帰分析の入門

分散と共分散

一何かを測るには具体的に数字として出る尺度が望ましい。単一の変数の変動を表す尺度が分散で、二つの変数の間の変動を表す尺度が共分散である。分散はゼロから無限大までの範囲で値が存在するが、共分散はマイナスの領域も値としてあり得る。また、共分散がマイナスだったら、二つの変数は負の相関を持ち、プラスだったら、正の相関を持つ。分散と共分散を利用すると、相関係数の計算ができる。

単回帰モデル

一二つの変数の間の線形的な関係性を表すモデルが単回帰モデルである。切片と回帰係数がパラメータであり、誤差項がモデル内に存在している。式の左に来る変数が被説明変数、右に来る変数が説明変数であり、説明変数が原因で被説明変数が結果の因果関係の解釈もできる。GPAと勉強時間、肺がん死亡者数とたばこ消費量など、様々な関係性を記述できる。さらに、回帰係数から具体的にどれくらいの変化が予測されるかを計算できる。

誤差

一数字として書く計測値にはすべて誤差が存在する。誤差には、その原因によって偶然誤差と系統誤差が分かれる。偶然誤差は全くの偶然が重なって生じる誤差で、回帰モデルの誤差項は偶然誤差をモデル化したものである。系統誤差は環境やモデルの間違いによる偶然ではない誤差であり、5分遅れている時計が測る時間などがそ

れに値する。

重回帰モデル

—重回帰モデルは共変量が2つ以上である線形関係を表す回帰モデルである。共変量はともに被説明変数の原因と思われる変数のことを指す概念で、例えば、パソコンの価格の決定要因にはCPUの価格、グラボの価格、マウスの価格などの多数の共変量が含まれる。重回帰モデルによって、様々な共変量が同時に変動した場合の被説明変数の予測値が分かる。

セテリス・パリブス

—セテリス・パリブスとは、他の変数を固定して、一つだけを変動させるという概念で、経済学における因果関係の議論によく用いられる。セテリス・パリブスが成り立っていれば、回帰モデルの分析結果を因果関係として解釈できる。

予想される問題

- A. 単回帰モデル、もしくは、重回帰モデルを提示して、その解釈を問う問題。
- B. 誤差の種類、偶然誤差と系統誤差の違いの理解度確認。
- C. セテリス・パリブスの概念から回帰モデルを分析する問題。

5. 回帰分析の例

例を扱う章は、試験の対象としない。

6. 回帰分析の罫

第6回小テストの内容の範囲で提出するので、第6回小テストの解説を中心に勉強。