

## 第1回 実験計画学とは？

### A. 授業の枠組み

#### 1. 講義について

実験計画学は統計学の一分野である。その理論そのものはあまりおもしろいとは感じられないだろう。理論を実際のデータに当てはめていくことで統計的な考え方を身につけるようになるのが望ましい。したがって、自分でデータを集め、そのデータを使った計算を必ずこなし、授業に積極的に参加して、統計的な考え方を育てることが必要である。

毎回、できるだけたくさんの学生に質問する。

ノートパソコン（授業ではエクセルおよびエクセルに付属する分析ツールの使い方を説明する）、これまでに配ったプリント（あるいはホームページ上にあるファイルをパソコンに入れておいてもよい）を持参のこと

授業のホームページ <http://www.ipc.shimane-u.ac.jp/food/kobayasi/>  
に資料などをのせておく。

#### 2. 宿題について

次回につながる宿題を出す。レポートの締め切りは原則として**翌週の月曜日午後1時**

配ったプリントは毎回必ず持ってくること

レポートはノートに書くか、プリントアウトをノートに貼るか、バインダーやクリアファイルなどに綴じて、データの入ったエクセルファイルを保存したフロッピーディスクかCDを添えて、提出する。ファイル名は実験計画学\_\_氏名とする（例：実験計画学\_\_小泉純一郎）。各回の宿題はそれぞれエクセルファイルに新しいシートを追加し、データを入力すること。レポートには毎回継続的な課題があるので、これまでの宿題も一緒に提出する。

授業の前にレポートを返却する。再提出の指示があるときは、次回のレポートとあわせて、再提出のレポートを提出すること。

#### 3. 成績について

以下の3つから成績を判定する。

- ① 授業参加 質問への受け答え、疑問点を質問したかなど
- ② レポート
- ③ 期末試験

## B. 経験から科学へ

### 1. 大数の法則

科学の始まり ケプラーの法則 ブラーエの集めた膨大な天文学データから  
それまでは（今でも？） 2, 3の事実（経験）でものをいっていた（いる？）  
根拠が薄い（柳の下のどじょう？）

ケプラーの法則から万有引力の法則へ

ニュートンはケプラーの法則から万有引力の法則を導いた  
万有引力の法則ですべての天体の運動を説明できる

多数のデータから少数の法則を導くことができた

海王星の存在の予言, ハレー彗星が周期的に地球の近くに来ることを予言など

たくさんデータを集めれば集めるほど確実である.

保険のデータ 人間の死や事故はわからないことだらけ

統計を集めて, 確率的に掛け金を決める (生命表 車の事故)

★ 統計をみて考えよう.

- ① 日本の出生性比はだいたい安定している. しかし, 明治 39 年と昭和 41 年には出生性比および出生率が前後の年に比べて大きく異なっている. なぜか?
  
- ② 人の寿命はそれぞれ異なる. しかし, たくさんのデータを集めると傾向がわかる. 日本人の平均余命 (2000 年), 年齢別・性別死亡率 (2000 年) をみて考えてみよう.
  
- ③ 40, 65, 80 歳まで生きる人の割合. 昭和 20 年代は定年 (65 歳, かつては 55 あるいは 60 歳だった) まで半数ぐらいしか生存しなかったが, いまや大半が生存する. 男でも半数は 80 歳まで生きられる.
  
- ④ 交通事故死者の年齢層別・状態別分類. 交通事故の実態を調べることによって, 自動車保険などの保険料を決めたり, 交通事故を防ぐための措置を考えたりする.

## 2. 少数例で何かいえないか？ 近代的な統計学の登場

大数の法則とはいっても、人はけちだからたった1回の経験で結論をつけたがる  
風邪が治った、米がたくさんとれた、ニワトリの卵が大きくなった  
タバコを吸うから長生きだという主張 自分だけがサンプルで結論する

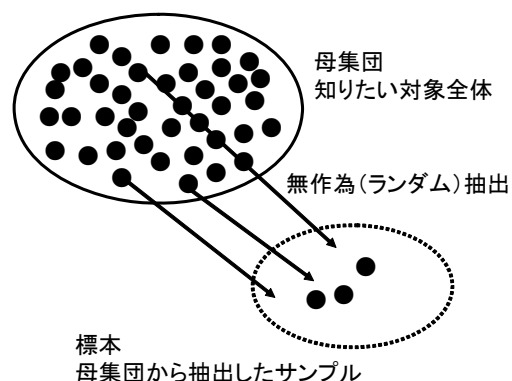
さらに大数の法則にかなうぐらいのデータを集めることは困難なことが多い  
自動車の耐久性テスト 少ない台数で、しかも正確なデータがほしい  
オオサンショウウオの食性を知りたい 数自体が少ない  
猫のエイズの治療薬 効かないかもしれない副作用の強い薬は動物虐待かも？

世の中には実験不可能なこともある

バブルの崩壊は誰のせい？ 世の中のことは1回限りである  
家庭環境と子供の人格形成 家庭環境は実験操作できない

それゆえ、できるだけ少ないデータでできるだけ多くの法則（結果）を得たい

標本（サンプル）から母集団を推定する 近代的な統計学の手法  
母集団（調査対象すべて）からランダム（無作為）に取ったサンプル（母集団からの代表）であれば、統計学からどの程度の精度・誤差で判断できるかがわかる。



正しく抽出した標本から母集団を推定する事例と誤差

視聴率 標本サイズと誤差 関東地区で 1455 万世帯から 600 世帯  
山陰地方で 45 万世帯 200 世帯をサンプリングする  
視聴率が 20%と出たら、その誤差はどのくらい？ どちらの精度が高い？

### 3. 正確なデータを取るために：標本の選び方は難しい

いい加減なデータ・統計からはいい加減なことしかいえない

統計の歴史

そもそも徴税あるいは徴兵の必要から 古代ローマ帝国, 古代中国, 太閤検地  
いい加減な統計からは悲劇しか出てこない 中国の大躍進

全部を調査することは不可能, あるいは時間, お金がかかりすぎる

国勢調査 全国民を調査する 5年に一度

標本に要求されること

母集団を代表しているか (統計的な手法を適用するための条件)

精度が必要な程度あるか (サンプルの数が多ければ精度は高くなる)

母集団を代表しているか インターネットを使って, 世論調査したら・・・?

元気そうなウシばかりサンプリングしたら?

無作為 (ランダム) なサンプリングであることが要請される

**無作為標本 (ランダムサンプル) ならば**

標本数が少数であっても母集団について代表値, 精度などを統計的に推定できる.

講義の第2, 3, 4, 5, 6, 7, 8, 12, 13回はこのようなランダムサンプルについての基本的な統計学を学ぶ.

### C. 実験計画学とは

大数の法則に従って, データをたくさん取ればより正確なことがわかるとはいえ, 実験を何度も繰り返すことはたいへん手間がかかる

したがって, 少ないサンプルで結論を得たい. しかし, サンプルが少ないと誤差が大きくなる・・・

#### 1. どのようにデータを取れば, 精度よく, しかも少ない実験ですむのか?

不均一なところ (圃場, 牛の集団) からどのように実験し, データを得るのか?

① 圃場によって地力が違うので, 同じ品種でも収量が異なる.

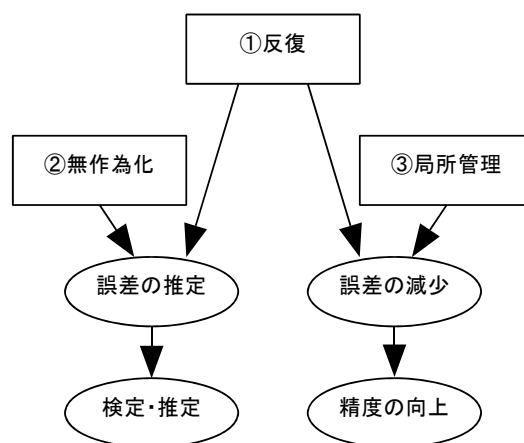
② ヒトに栄養剤を与えた効果を知りたいが, 体の大きさによって効果が違いそうだ.

では実験の場を人為的に均一にすればよいのか?

① 圃場の土を均一に混ぜてむらをなくした圃場 (精密圃場) の問題 多収品種の判定

② 同じ身長・体重のヒトだけを集めてよいのか?

- 不均一さの克服のためには
1. 精度を知る どの程度ばらついているのか?
  2. 傾向のある不均一さを偶然誤差に転化する
  3. 可能ならば不均一さによる誤差を除去して、精度を高める



## 2. 実験計画学の基本（フィッシャーの3原則）

- ① 反復—誤差分散の評価
- ② 無作為化—一定誤差の克服
- ③ 局所管理—一定誤差の除去（実験精度の向上）

### ① 反復

統計的検定・推定の基礎になる誤差の大きさを、その実験自身から評価するためには、同じ条件の下での2回以上の繰り返し（反復）が必要である。これによって、その実験から得られるデータのばらつきの程度が推定されるからである。

### ② 無作為化

実験で起こる誤差は大きく分けて、偶然誤差と系統誤差である。偶然誤差は原理的にも消去が不可能である。系統誤差は実験方法によって消去できるが、圃場の地力差、動物の個体間差など消去が困難あるいは不可能なものもある。さらに系統誤差の中にもあらかじめわからないものもある。このような系統誤差をサンプリングの無作為化によって、偶然誤差に転化させることによって、結果の判断に意図的な偏りが起こらないようにする。

### ③ 局所管理

ばらつきのあるものでも、ある程度ならばそのばらつきを小さく抑えることができる。例えば、圃場には地力差があるが、小さい区画であれば、その差を小さく抑えることができる。あるいは、体重に近い動物を集めることもできる。このように実験全体は大きなばらつきを持つところから代表を集めても、その中でばらつきを小さくして管理した実験を行うことができる。これを局所管理という。

## 3. 実験計画学の3つの柱

- ① 誤差の定量と制御（精度がよくなる・精度がわかる）
- ② 実験回数を少なくできる
- ③ 実験のデータの変動のうち、処理による意味のある部分と偶然誤差を分けることができる。  
さらに処理の主効果と交互作用を検出できる。

講義の第9-11, 14回は実験計画学の基本を学ぶ。

#### D. 宿題

1. エクセルで授業に必要な、基本的な計算ができるように、ホームページの練習問題を各自解けるようにしておく（レポートを提出する必要はない）。
2. 大数の法則が成り立つくらいのデータを集める。最低でも 100 以上のデータを集める。同じことを2人以上が調べるようなことがないように、自分が調べることを 204 室前のホワイトボードに貼った名簿に調査テーマを事前を書く（早い者勝ちとする）。なお調査テーマが大数の法則にはふさわしくないものであればテーマをもう一度考え直させる。承認を得ないテーマで調べても、レポートは受け取らない。

次のような条件を満たしているデータでなければならない。

- 1) 個々のデータには誤差が含まれている。
- 2) 2, 3 のデータではデータ全体の特徴はよくわからないが、データ数を増やしていくとデータ全体の特徴が次第に明らかになっていく

以下のようなものはだめである。

- 1) データ全体の個数が1万以下しかないもの  
世界各国のGDP, 市町村の人口, 株価など
- 2) データにばらつき（誤差）がないもの  
車のナンバープレートなどはそれぞれの値がそれぞれ固有の値であり、誤差がありえない。
- 3) 得られたデータが母集団を代表しないもの  
世界の山の高さなど・・・得られたデータは標高の高い山ばかりだから世界の山を代表しない
- 4) 大数の法則に従わない時系列のデータ  
1901年から2000年までの日本の人口, 上位100人の・・・など  
地震の回数などは大数の法則に従うのでよい
- 5) 名目データ, 順序データ, 順位データ  
色, 性別などのデータは, 以下の授業で計算する平均, 標準偏差などの計算をすることができない。

データの種類は大きく、5つに分けられる。

名目データ：順序、大きさのない属性データ。例えば、性別、色、血液型など。

順序データ：順序、段階だけを示したデータ。大、中、小あるいは数字で1～5（成績）など。

順位データ：順位を付けたデータ。

離散データ：とびとびの値をとるデータ。人数、個数、値段など。

連続データ：連続的な値を取るデータ。身長、速度、濃度など。

調査例：イネの穂の長さ（連続データ）

ニワトリの卵の重さ（連続データ）

大橋川に生えているヨシの草丈（連続データ）

司馬遼太郎の小説に使われる漢字の画数の分析（離散データ）

宍道湖のシジミの収穫個数（離散データ）

松江のスーパーで売られる豚肉の値段（離散データ）

1. 何か法則性がありそうなデータを集める。データを集める前に、どんな法則が出てくるかを予想する。
2. 100以上のデータを集め、大きさの順に並べる。
3. 次の値を見つけだすあるいは計算する。
  - 最大値、最小値、レンジ（最大値と最小値の差）
  - 平均
  - メジアン（中央値：大きい順に並べたときまん中に来る値。データが偶数のときはまん中に位置する2つの値の平均）
4. 分布をグラフにする。
  - 最大値と最小値の間を5, 10, 15に分級してそれぞれ頻度分布を書く。
5. ここまでの時点で観察の結果、わかったことを、最初の予想と照らし合わせながら書く。箇条書きでよいから、なるべくたくさん書くのがよい。なお上の2.～4.の手順に基づいてデータをまとめていくとき、どの手順の時にわかったかをなるべく書くこと。

宿題の見本はホームページ <http://www.ipc.shimane-u.ac.jp/food/kobayasi/> を見てください。

宿題はノートに書くか、あるいはプリントアウトをノートに貼るか、バインダーやクリアファイルなどに入れて、データの入ったエクセルファイルを保存したフロッピーディスクかCDを添えて、11日（火）午前9時までに生物資源科学部2号館204室に提出のこと。（メールでの提出は認めない）

実験計画学予定表

1. 10月 4日 実験計画学とは何か?
2. 10月11日 平均と分散・データの要約
3. 10月18日 母集団と標本, 確率分布
4. 10月25日 二項分布, ポアソン分布, 正規分布
5. 11月 1日 統計的推定
6. 11月 8日 統計的検定
7. 11月15日 t分布とt検定
8. 11月22日 カイ二乗分布, F分布とその応用
9. 11月29日 分散分析その1 一元配置
10. 12月 6日 分散分析その2 二元配置
11. 12月13日 分散分析その3 実験計画法
12. 12月20日 相関分析
13. 1月17日 回帰分析
14. 1月24日 実験計画法と統計解析の実際
15. 2月 7日 期末試験 (期日変更の可能性が大きい)