



# 確率と統計

林クラス  
第3週

SA 荒木 楽平 小暮 達也

# 第1回レポート課題

締切: 4月30日(金)12:00(長い連休前に出そう!)

提出場所: 11号館集合教員室前レポート提出箱No.8

内容: 以下を行い, 別紙で作成した表紙とともに1つにまとめて左上をホチキスで留め, 提出せよ.

1. 第2章 練習問題(1)~(3)を行い, 実行状況を印刷する. **各問題の作業状況を赤ペン等で囲んで明確にし**, 対応する問題番号を書き込むこと. 作成したグラフも印刷すること.
2. 練習問題を行った結果を用い, 2つの大学を比べて分かることを述べよ. **手書きでレポート用紙に記入すること**. PCでの作成は認めない.

# 第1回レポート課題続き

用紙サイズ:A4

表紙をレポート内容とは別の用紙で作成し、十分大きな文字で表題、名前、クラス・名列番号を記入すること。

表題:確率と統計 第1回レポート

レポートを作成するためのいろいろなヒントを以下のスライドにまとめた。参考にすること。

授業時間中に一通り試してみて、うまく行かない場合は申し出ること。

# Rの作業内容を印刷する

RのGUI画面では左上にアイコン列がある。

一番右のプリンタアイコンをクリックすると印刷できる。

- 作業状況のウィンドウをクリックしてプリンタアイコンをクリックする→作業状況が印刷できる。

**注意:** 失敗した作業内容が含まれていても、大量でない場合はOKである。(うまくいった部分を赤ペンなどで囲んで明記することになっているため。)

- グラフ表示のウィンドウをクリックしてプリンタアイコンをクリックする→グラフを印刷できる。

今回のレポートでは、以上のやり方でOK。

# トラブルシューティング

- 授業中など、プリンタがその場にはない場合  
→ 次のスライド参照. 印刷時にpdfファイルに書き出せるが、林自身はそのpdfファイルを開こうとするとエラーになってしまい、うまくいっていない.
- Rは使い慣れないので、入力ミスだらけになった！  
→ あまりに大量になる場合は、次のスライドを参考にしてやり直して欲しいが、うまくいった箇所を蛍光ペン等で囲み、問題番号と対応付けしてあれば良い. 目安はうまくいった箇所が全体の半分程度あること.

その他、レポートに関する質問はこの時間中にすること！

# Rで作業するためのTips

- 起動時のワークスペース設定→教科書P.394にやり方がある. この付録CはFAQなので, 読んでおくが良い.
- レポートのために作業状況を保存する方法の1つ
  1. 「ファイル」メニューで「ファイルを保存...」を選ぶ.
  2. ファイル名などを指定するウィンドウが開くので, テキストファイルとして保存する.
  3. 保存したファイルをメモ帳で開くことができる.
    - そのファイルを印刷すると作業状況が印刷できる.  
→レポート課題作成
    - そのファイルからうまくいったコマンドをコピー&ペーストすると便利そう.

# Rの作業スペースとは

Rの「作業スペース」: それまで計算して代入した変数とその値 **注意！それまでの作業そのままではない.**

Rの終了前に「作業スペースを保存しますか？」と尋ねられ、保存すると次回はその作業スペースを自動的に読み込む.

→ 毎回作業スペースを保存しつづけると、どんどん変数などが増えていくことに注意する.

確認: 例えば `c<-3+5` を実行し、Rを終了する. このとき作業スペースを保存しておく.

→ Rを起動し、すぐに `c`(Enter)してみよう. 計算していないはずなのに、8が表示される.

# 今日の内容

教科書第2章 1つの変数の記述統計 後半

2.7 散布度: データの散らばり方(終わり)

2.8 分散と標準偏差

2.9 散布度を表す他の指標

2.10 標準化

2.11 偏差値

教科書に書いてあることを実際自分の手で打ち込み  
ながら, Rで各種統計量を計算してみよう.

(動作を確認したところ, 問題はないようであった.)

## 2.8のポイントと補足

分散には「(各データと平均の間の2乗距離)の平均」という意味がある.

→散らばりが大きければ分散は大きく, 散らばりが小さければ分散は小さくなる.

→散らばり方を表す指標になる.

分散はデータの2乗の大きさになる

→データとそろえるために分散の平方根を取る  
=標準偏差

## 2.8のポイントと補足(続き)

分散には2種類ある。→標本分散と不偏分散

$$V = \frac{\sum_{i=1}^n (d_i - a)^2}{k}$$

標本分散: 分母kがデータ個数

データ自体のちらばりを示す場合に使う。

不偏分散: 分母kが(データ個数-1)

データから母集団の値を推測する場合に使う。

Rのvar関数は不偏分散なので、標本分散を計算したい場合は(varの結果)×(データ数-1)／(データ数)で計算する。

## 2.9のポイント

散布度の指標に一番使われるのは分散と標準偏差  
それ以外の散布度の指標には以下のものがある.

平均偏差:

(データー平均)の絶対値の総和 / データ個数

Rで絶対値を計算するのはabs関数

範囲: データ中の最大値 - データ中の最小値

Rで最大値を計算するのはmax関数

最小値

min関数

## 2.10のポイント

標準化とは: 平均0, 標準偏差1になるように全てのデータの値を同じ式で変換する.

→変換後の値がz得点

例えば, 2回の試験があって, それぞれの試験の

- 自分の得点
- 全体の平均点
- 全体の分散(標準偏差)

がわかっているならば, 全データを持っていなくても, 順位の変動を計算できる→偏差値

## 2.11のポイント

偏差値：平均50，標準偏差10になるように標準化した  
得点.

→偏差値 =  $Z$ 得点  $\times 10 + 50$

一番身近な使用例は，高校入試，大学入試のための模  
擬テスト.

試験ごとに平均点(難易度)が違う.

→偏差値で，全体の中で自分がどのくらいの順位なの  
かがわかる.

# 次回の予定

4月29日(木, 休): 昭和の日(講義なし)

5月6日(木): 第3章 2つの変数の記述統計

第3章の練習問題を第2回レポートとして出題する予定  
である.