

# 統計学第 1 講/第 2 講

後藤 晶

akiragoto@meiji.ac.jp

## 目次

第 1 講：イントロダクション	2
教員の自己紹介	2
講義の概要	2
授業で使うアプリケーションの紹介	6
統計学やデータサイエンスを学ぶ意義を考える	6
第 2 講：基本的な操作法と記述統計量の算出	7
R を使って計算しよう	7
実際に計算をしてみよう	9
記述統計量とは	10
分散と標準偏差を手計算で算出してみよう	15
分子を計算する	15
分母を計算する	16
最後の計算	17
標準偏差を算出する	17
2 種類の分散と標準偏差	18
今日の課題	19
おうちでやっておこう	20

次回の案内：	20
R でデータを扱う時に注意すべきこと	20

## 第 1 講：イントロダクション

### 教員の自己紹介

#### 軽く自己紹介を...

- 名前：後藤晶（ごとうあきら）
- 1984 年 7 月 13 日生まれ
  - － 35 歳，独身＞＜；
- 出身地：横浜
  - － いわゆる「ハマっ子」です．
  - － 高校は横浜駅から歩いて通ってました．

#### 軽く自己紹介を...

- 趣味：読書，音楽聞いたり，一人旅
  - － イロイロ読みます．
  - － 割と激し目の音楽が好き
- 専門分野：行動経済学・社会情報学・実験/計算社会科学
  - － 自発的貢献行動（協力行動）の促進／抑制要因：ゲーム理論に関する実験を中心として
  - － 行動経済学の観点からの「政策」評価：行動経済学を政策に活かさないか？
  - － クラウドソーシングを用いた経済ゲーム実験環境の構築

## 講義の概要

### 授業概要

本講義では仮説をデータに基づいて統計的に検証したり、データ解析の結果から、新たな事実を発見したりするときに役立つ統計的手法を身に付ける。

講義はデータの整理の仕方、平均、分散等を求める記述統計学より始め、確率、母集団、標本抽出、確率分布を学び、最終的に推定、検定といった推測統計学を解説、演習する。本講義は、理学療法学研究法、作業療法研究法、卒業研究の基礎となる科目である。

## ざっくりいうと

- データとの付き合い方を学ぶために、R を使って分析手法を学ぶ。
  - R を使いやすくするために RStudio を使用する。
- 基本的なデータの扱い方を整理して、グラフの可視化などを学ぶ。
- 分析手法として、回帰分析、t 検定、分散分析、重回帰分析を学ぶ。
- 最後はグループプレゼンテーションを評価する。
  - もしくはレポートなど、個人での演習となる場合もある。
  - 様子を見ながら、プレゼンテーション資料をもとにしたレポート資料についても評価する。

## 到達目標

1. 各統計手法について、その目的と意義を説明することができる。
2. 各統計手法について、各自で分析を実行できる。
3. 分析結果について、適切に他者に説明できる。

## 授業内容

- 【第 1 講】イントロダクション
- 【第 2 講】基本的な操作法と記述統計量の算出
- 【第 3 講】データの可視化
- 【第 4 講】実証分析の手続き
- 【第 5 講】カテゴリーデータの分析
- 【第 6 講】カテゴリーデータの分析
- 【第 7 講】差の検定
- 【第 8 講】差の検定

## 授業内容

- 【第 9 講】相関分析・回帰分析
- 【第 10 講】相関分析・回帰分析
- 【第 11 講】一要因分散分析
- 【第 12 講】一要因分散分析
- 【第 13 講】重回帰分析
- 【第 14 講】重回帰分析
- 【第 15 講】発表

※ ただし、履修者の状況により内容を一部変更することがある。

## 評価

- 平常点：30%
  - － 毎回の授業での取組状況を評価します。
- 講義内の課題：50%
  - － 毎回の講義の中で演習問題を出します。こちらの課題を提出してください。
- 演習問題：20%
  - － 最後に演習問題を課します。今のところ、グループプレゼンテーションを予定していますが、社会環境の状況に応じて、各自での演習問題となることがあります。

## 課題に対するフィードバックの方法

- フィードバックとして、課題に対して全体的なコメントを返します。
  - － 具体的には、前の講義での演習問題について、次の講義でコメントをします。

## 教科書

- 必要な資料をオンラインで配布します。
  - － 配布 URL は以下のとおりです。
  - － <https://akrgt.github.io/2020statistics/>

## 参考図書

- Navarro, D.J, & D.R.Foxcroft（著）芝田征司（訳），『jamovi で学ぶ心理統計』

## 統計について

- 小杉考司, 2019, 『言葉と数式で理解する多変量解析入門』, 北大路出版
  - － 非常に平易な言葉で多変量解析について説明がなされている。ただし、一歩戻って基本的 RStudio の使い方と基本的な統計 (t 検定や回帰分析, カイ 2 乗検定など) を学ぶのであれば, 小杉考司, 2019, 『R でらくらく心理統計 RStudio 徹底活用』などもよいでしょう。
- 星野匡郎, 田中久稔, 2018, 『R による実証分析—回帰分析から因果分析へ—』, 裳華房
  - － R を用いた様々な分析手法について記載されている。春学期の内容も含まれるが, 主に秋学期の内容をカバーしている。
- 川端一光, 岩間徳兼, 鈴木雅之, 2018, 『R による多変量解析入門 - データ解析の実践と理論』, オーム社
  - － 基本的な技術を学ぶことができる。かなり良書。
- 森田果, 2014, 『実証分析入門 - データから「因果関係」を読み解く作法』, 日本評論社

- 法律系の先生が書かれた本だが，なかなかわかりやすく面白い．どちらかというと秋学期の内容に関わる．
- 地道正行, 2018, 『データサイエンスの基礎 R による統計学独習』, 裳華房
  - 細かく書かれているが，数式も多い．講義内で数学的説明の時間を十分に確保できないため，興味のある方はこちらへ．

## 再現性の議論について

- 高橋康介, 2018, 『再現可能性のすゝめ (Wonderful R 3)』, 共立出版
- 江口哲史 (編), 2018 『自然科学研究のための R 入門—再現可能なレポート執筆実践—(Wonderful R 4)』, 共立出版
  - いずれの本も再現可能性のために RStudio と RMarkdown の使い方について紹介した本．

## 分析の一連の流れについて

- 松村優哉, 湯谷啓明, 紀ノ定保礼, 前田和寛, 2018, 『R ユーザのための RStudio[実践] 入門—tidyverse によるモダンな分析フローの世界』, 技術評論社
  - RStudio を用いたデータ収集, データ整形, 可視化, レポートニングといった一連の分析の流れに関する本．

## 参考 web 資料

- からだにいいもの
  - R に関する様々な情報が掲載されている．多少応用的なトピックが多い．
- marketechlabo
  - ちょっと新しいパッケージ等が紹介されていて興味深い

## アクティブ・ラーニング

- グループワークおよびプレゼンテーションを行う．

## 留意事項

- R, RStudio という統計ソフトを使って，実践的に分析を進めながら統計学を学んでいきます．
- 各自が参考図書等に目を通してながら，積極的に学ばれることを期待します．可能であれば，自宅の PC にもインストールをしてください．
- インストール方法は，簡単に動画でまとめてあるので自宅に帰ってからご確認ください．

## 授業で使うアプリケーションの紹介

### 授業で使うアプリケーション

- R & RStudio：基本的な分析手法から高度な分析手法まで、必要な分析は何でもできる。
  - しかもフリーと来たもんだ！

### この授業でフリーのアプリケーションを使う理由：

- 様々な有料のアプリケーションは存在する。
  - SPSS, Stata, SAS, Eviews, などなど...
  - 企業に就職してもこれらのアプリケーションが使えるとは限らない。
  - 家で勉強しようにも有料で困る＞＜；
- フリーソフトであれば、いつでもどこでも必要に応じて自分で分析ができる。
  - 卒論に限らず、就職してからでも使えるスキルとして身につけられる。
  - そういった観点から、この授業では R を中心としたフリーソフトを使います。

## 統計学やデータサイエンスを学ぶ意義を考える

### 統計学とは？

- データサイエンス：データから有用な情報・知識を引き出したり、新たな価値あるデータを創造するための基本的な考え方
  - 平均・分散を算出したり、全体的な傾向を把握するためにヒストグラムを作るのは有用な情報・知識を引き出すため。
  - プログラムを組んで、新たに価値あるものを作る。
- 統計学を学ぶと何がどうなる？
  - 「データ」に基づいた思考方法が身につく
  - 「見せかけの類似性」に騙されなくなる。
  - 日常生活・ビジネスへの応用可能性が広がる

### 「血液型占い」：血液型によって性格が異なる。

- 「データ」に基づくと、血液型によって性格が異なるとはいえない（参考）。

## 見せかけの類似性：相関関係と因果関係.

- 「データ」はモノを考えるのに大事なことだが、「数字」や「見た目」に騙されてはいけない.
- その他の要因が影響している可能性がある（参考）.

## 日常生活・ビジネスへの応用可能性が広がる

- ビジネスにおいて、データを利用し、企業活動を改善・開拓するのに必要な 3 要素
- データを効率的に収集・処理する
  - 企業活動におけるデータは莫大であり、このようなデータを処理するためには情報技術の利用が必要
- データを適切に取り扱い、妥当かつ汎用的な成果を残す
  - データ全体から、その傾向の妥当性を検討する.
- データをビジネスの枠組みの中うまく組み込む
  - 効率的に処理をし、有用な結論を導き出してビジネスへの応用をしていく.

## 第 2 講：基本的な操作法と記述統計量の算出

### R を使って計算しよう

#### 今回の目標：

- RStudio の準備をしよう.
- R を使った基本的な計算方法に慣れよう.
- 分散・標準偏差を手計算で算出してみよう.

#### RStudio を使うための準備

- RStudio で使うディレクトリを決める
  - 初回のみ準備が必要
- “.R” ファイルを作成する
  - 毎回準備が必要

#### RStudio で使うディレクトリを決める

- 最初に RStudio で使うディレクトリ（フォルダ）を決めます. この時, 【フォルダまでの間に 2 バイト文字（日本語）やスペースが入らないように気をつけて下さい】.
  - 使えない文字：あいうえおかきくけこ「」[] 今日の天気は晴れ！1 2 3 4 5 6 7 8 9 0

- 使える文字 : abcdefghijklmnopqrstuvwxyz ABCDEFGHIJKLMNOPQRSTUVWXYZ1234567890
- また, windows でも mac でもこの作業は変わりません.

## “.Rproj” ファイルを作成する

- 毎回, 授業の際に R で作業をする際には, 指定した作業用フォルダにある “.Rproj” ファイルから開きます.
- 作業フォルダを決める (必要に応じてフォルダを作成する)
- RStudio を起動する.
- 右上の現在いるフォルダを示している文字をクリックする.
  - 場合によっては “(None)” と表示されているかもしれません.
- “New Project...” をクリックする.
- “Existing Directory” をクリックする.
- “Browse...” をクリックして, 先程決めた作業フォルダまでたどり着く
- “Open” をクリックする.
- “Create Project” をクリックする.
- 完了

## “.R” ファイルを作成する

- 授業中に作成したファイルを保存するために, “.R” ファイルを作成する.
  - これは毎回行います.
- 左上にある白い四角の左上に緑のプラスが書いてあるヤツをクリックする.
- “R Script” をクリックする.
- “Untitled 1” と書かれたファイルができる.
- “ctrl + s” を押すと保存ができる.
  - Mac で作業している場合は “cmd + s”
- ファイル名として “今日の日付 \_ 学籍番号” を設定する.
  - ex: “20200929 \_ 学籍番号.R” とする
  - 学籍番号を入れてもらうのは, 後ほど提出してもらうため.



- “Save” をクリックする。
- また、このファイルに書いた数式は“ctrl + Enter”（Mac の場合は“cmd + Enter”）でその行の計算を R に読み込ませることができます。

## 実際に計算を試みよう

### 加減乗除

```
123 + 123 # 足し算
```

```
## [1] 246
```

```
123 - 123 # 引き算
```

```
## [1] 0
```

```
23 * 123 # 掛け算
```

```
## [1] 2829
```

```
123 / 123 # 割り算
```

```
## [1] 1
```

```
123 ^ 2 # 累乗
```

```
## [1] 15129
```

加減乗除の基本はこのような形です。少し演習問題を解いてみましょう。

### 注意です。

- 「#」から始まるとコメントアウトができます。
- 要は、「計算式」ではなく、「ただの文字」として認識されます。

### 演習問題 1

- $113 + 987$
- $2135 + 231$
- $9832 - 3422$
- $12348 - 8976$
- $17 * 16$
- $3298 * 5$

- $285195 / 5$
- $12387 * 33$
- $324 ^ 2$
- $89 ^ 4$

## その他の計算

```
sqrt(144)# 平方根
```

```
## [1] 12
```

```
1234 %/% 123 # 整数商：割り算をした時の整数部分
```

```
## [1] 10
```

```
1234 %% 123 # 剰余：割り算のあまり
```

```
## [1] 4
```

## 演習問題 2

- $(34 \times 2) + (43 - 12)$
- $23 \times (92 - 9)$
- $(53 + 23)$  の 5 乗
- $(334 - 56) \div 90$
- $(34 \times 2) + (43 - 12)$
- $(3221 + 239) \div (87 + 27)$  の整数商
- $(751 \times 90) \div (5412 / 32)$  の剰余

## 記述統計量とは

### 平均値・分散・標準偏差とは？

- 平均値：全てのデータを足して割ったもの。一般的に代表値（データ全体を表している数値）として扱われる。
- 分散：平均値とそれぞれの値の差を求めて 2 乗して、合計したものをデータの個数で割ったもの。データの散らばり具合を示す数値であり、分散が大きければ大きいほど、データが散らばっていることを示す。
  - $\sigma^2$  という記号で表される。
  - $( ) = \Sigma \{ ( ) - ( ) \}^2 / ( )$
- 標準偏差：分散の平方根。通常の長さのばらつきを評価する際には同じ単位で理解したほうがわかりや

すいために用いる。

- $\sigma$  という記号で表される。

## その他、重要な指標

- 最小値：そのデータの中で最も小さい値
- 第一四分位数（25% パーセンタイル値）：最小値と中央値の間の中央値
- 中央値（第二四分位数）：データを大きい（小さい）順に並べたとき、真ん中の値のこと（median）。外れ値がある時に代表値として用いられる。
  - 奇数の場合：ちょうど真ん中が存在する。
  - 偶数の場合：真ん中の数字 2 つの平均値を中央値とする。
- 最頻値：データの中で最も多く出てくる値のこと（mode）。因子データの際に代表値として使われる。
- 第三四分位数（75% パーセンタイル値）：中央値と最大値の間の中央値
- 最大値：そのデータの中で最も大きい数
- 以下の 2 つは参考までに。
  - 平均偏差：「平均からの偏差」の絶対値の平均
  - 範囲：最大値から最小値の間。引き算で求められる。

## 平均値の計算

- 7 人の学生の体重が 50, 60, 85, 70, 80, 67, 66kg であったとする。これらの学生の体重の平均値を求めよ。
- 平均値：全てのデータを足して割ったもの。一般的に代表値（データ全体を表している数値）として扱われる。

## 平均値の計算

- 7 人の学生の体重が 50, 60, 85, 70, 80, 67, 66kg であったとする。これらの学生の体重の平均値を求めよ。

```
# 平均値 =(それぞれのデータの値の合計)/(データの個数)
(50+60+85+70+80+67+66)/7
```

```
## [1] 68.28571
```

## オブジェクト指向

- 「オブジェクト」：データやモデル式などを入れる「何でも箱」
  - R ではモデル式、データなどをオブジェクトに入れて考える
  - 数式やデータをいちいち書くのは大変...
  - オブジェクトに入れることを「代入する」と言う

オブジェクトに入れて計算する：

```
x <- 5 #x というオブジェクトに 5 を代入する
x #x の値を出力する
```

```
## [1] 5
```

```
(y <- 3) #( ) に挟むと、一発で結果も出してくれる。
```

```
## [1] 3
```

```
x + y
```

```
## [1] 8
```

```
x * y
```

```
## [1] 15
```

```
x - y
```

```
## [1] 2
```

よくないオブジェクト：

```
x <- 5
x <- 16
x
```

```
## [1] 16
```

- これで出力すると、x に 16 が代入されてしまっている。
- 基本的には違う語を使うようにしたい。

```
# NA <- 3
```

「NA <- 3 でエラー：代入の左辺が不正 (*do\_set*) です」と怒られる。これは「NA」がデータがないことを示す記号として指定されているため。他にも指定されている語がいくつかあるが、怒られたら違う文字を割り振れば良い。

オブジェクトには文字を入れることが可能

```
A <- "Pen"
B <- "Pineapple"
```

```
C <- "Apple"
paste(A, B, C, A)
```

```
## [1] "Pen Pineapple Apple Pen"
```

- paste : 複数の文字列を結合して、一つの文字列にする関数

## データセットを作ろう

7 人の学生の体重が 50, 60, 85, 70, 80, 67, 66kg であったとする。このデータを変数名 “weight” に代入する。

```
weight<-c(50, 60, 85, 70, 80, 67, 66)
```

### 演習問題 3 :

同じ 7 人の学生の身長が 155, 164, 182, 165, 177, 177, 172cm であったとする。このデータを変数名 “height” に代入せよ。

## R における「関数」とは？

- 関数 : 頻繁に用いられるデータ操作方法や、標準的な統計計算をまとめてオブジェクトにしたもの。正式には「関数オブジェクト」
  - 簡単に計算できるように、先人たちがまとめたものだとして理解すれば良い。
  - これを用いることで、簡単に計算ができる。

## 基本的な関数 :

- sum : 合計
- mean : 平均値
- max : 最大値
- min : 最小値
- range : 範囲 (最大値-最小値)
- median : 中央値
- var : 不偏分散
- sd : 標準偏差
- quantile : 四分位点
- IQR : 四分位範囲
- summary : 要約統計量
- sqrt : 平方根

基本的な関数：

- abs：絶対値
- round：値の丸め
- floor：値の切り捨て
- ceiling：値の切り上げ
- log：自然対数
- log10：10 を底とする対数
- log2：2 を底とする対数
- log1p：1 を加算した自然対数
- exp：指数関数
- sin, cos, tan：三角関数
- asin, acos, atan：三角関数の逆関数

記述統計量を色々出してみる.

```
sum(weight)/7  #sum() 関数で合計を算出できる.
```

```
## [1] 68.28571
```

```
sum(weight)/length(weight)  #length() 関数でデータの個数を数える.
```

```
## [1] 68.28571
```

```
mean(weight)  #実は mean() という関数を使うと一発で出てしまう.
```

```
## [1] 68.28571
```

```
median(weight)  #中央値は median() という関数で出せる.
```

```
## [1] 67
```

```
table(weight)  #最頻値は table() という関数を使って探し出す.
```

```
## weight
```

```
## 50 60 66 67 70 80 85
```

```
## 1 1 1 1 1 1 1
```

```
# ちなみに, "weight" の中に最頻値は存在していない. (全てが最頻値 =1)
```

#### 演習問題 4 :

変数名 “height” の合計・個数・平均値・中央値・最頻値を求めよ.

体重の記述統計量をまとめて算出する.

```
summary(weight)
```

```
##      Min. 1st Qu.  Median     Mean 3rd Qu.     Max.
##   50.00   63.00   67.00   68.29   75.00   85.00
```

- 左から順番に「最小値, 第 1 四分位数, 中央値, 平均値, 第 3 四分位数, 最大値」を示しています.

#### 演習問題 5 :

変数名 “height” の最小値・第 1 四分位数・中央値・平均値・第 3 四分位数・最大値を求めよ.

## 分散と標準偏差を手計算で算出してみよう

### 分散を算出する

- 定義通りの算出方法

$$\sigma^2 = \Sigma(( ) - ( ))^2 / ( )$$

- 簡便に算出する際に用いられる数式

$$\sigma^2 = \Sigma(( )^2 / ( )) - (\Sigma(( ) / ( ))^2)$$

\* 「2 乗の平均」 - 「平均の 2 乗」

- ここでは, 定義通りの数式から分母と分子に分けて話を進めていきましょう.

### 分子を計算する

#### 体重の平均値をオブジェクトに入れる

- “mean\_weight” というオブジェクトを作って, 体重の平均値を入れます.

```
mean_weight <- mean(weight)
mean_weight
```

```
## [1] 68.28571
```

平均からの偏差を求めて、オブジェクトに入れる

\*(データの値)-(weight の平均値) をして「平均からの偏差」を求めます。結果は“hensa\_weight”に代入します。

```
hensa_weight <- weight - mean_weight
hensa_weight
```

```
## [1] -18.285714 -8.285714 16.714286 1.714286 11.714286 -1.285714 -2.285714
```

「平均からの偏差」を 2 乗する

\*「平均からの偏差」を 2 乗します。“hensa\_weight2”というオブジェクトを作って代入をしましょう。- 2 乗しないと全部足すと、数字は 0 になります。- ただし、小数点以下を四捨五入しているので、ここでは完璧に 0 にはなりません、限りなく 0 に近くなります。

```
hensa_weight2 <- hensa_weight^2
hensa_weight2
```

```
## [1] 334.367347 68.653061 279.367347 2.938776 137.224490 1.653061 5.224490
```

「平均からの偏差の 2 乗」を全部足してオブジェクトに入れる

- これらの 5 つの値を合計した「平均からの偏差の二乗和」を求めます。
  - “sum\_hensa\_weight2”という名前にしましょう。これで分子は完成です。

```
sum_hensa_weight2<-sum(hensa_weight2)
sum_hensa_weight2
```

```
## [1] 829.4286
```

分母を計算する

データの個数を数えてオブジェクトに入れる

- 今度は分母を算出します。分母はデータ数です，“length\_weight”というオブジェクトに代入しましょう。



```
length_weight<-length(weight)
length_weight
```

```
## [1] 7
```

## 最後の計算

### 分散の算出

- 分散は「平均からの偏差の二乗和」 / 「データ数」ですから、以下の通りに求められます。
  - 分散は “vari\_weight” というオブジェクトに入れましょう

```
vari_weight<-sum_hensa_weight2/length_weight
vari_weight
```

```
## [1] 118.4898
```

## 標準偏差を算出する

### 標準偏差を算出する

$$\sigma^2 = \Sigma(( )^2 / ( )) - (\Sigma(( ) / ( ))^2)$$

$$\sigma = \sqrt{\sigma^2}$$

- 標準偏差は分散の平方根です。
  - 平方根を求める関数は “sqrt( )” であり, “hyohen\_weight” というオブジェクトに入れてあげます.

### ルートをとります

```
hyohen_weight <- sqrt(vari_weight)
hyohen_weight
```

```
## [1] 10.8853
```

### 分散の別解

$$\sigma^2 = \Sigma(( )^2 / ( )) - (\Sigma(( ) / ( ))^2)$$

```
sum(weight^2/length(weight))-sum(mean(weight)^2)
```

```
## [1] 118.4898
```

## 演習問題 6 :

- “weight”と同様に，変数名 “height” について分散・標準偏差を計算してください.
  - ヒント：ちょっと書き換えるだけですぐいけます.
  - “ctrl + f”(mac では “cmd + f”) で置換ができます.
  - 基本方針は「極力手抜きをしましょう」です.

関数を使って分散と標準偏差を算出する.

- 分散と標準偏差はよく算出します. 当然 R では関数が用意されています.

```
var(weight) # 分散
```

```
## [1] 138.2381
```

```
sd(weight) # 標準偏差
```

```
## [1] 11.75747
```

- 関数で求められるのは「不偏分散・不偏標準偏差」
- 手計算で求めたのは「標本分散・標本標準偏差」

## 2 種類の分散と標準偏差

### 2 種類の分散と標準偏差 ??

「不偏分散・不偏標準偏差」と「標本分散・標本標準偏差」というものが出てきました. この話を理解するためには「母集団」と「標本」という話を理解する必要があります. ここでは簡単に, その 2 つの違いについてお話ししたいと思います.

私達が何かのデータを取る時は, 全ての物事のデータを集めることが必ずしもできるとは限りません. 例えば, 「本学大学生 1 年生全員を対象としたアンケート」を実施すれば全てのデータを集めることができるかもしれませんが, 「日本国民全てを対象としたアンケート」を集計するのは非常に困難です.

例えば, 大学 1 年生の意見を調査することを目的として, 1 年生全員のデータをそのまま用いる分には問題ないのですが, 「日本国民全てを対象としたアンケート」を実施するのはコストの面から考えても現実的ではありません. そのために, 全体 (母集団) の中から一部を取り出して (標本, サンプル), 全体の意見・傾向を「推定」という手法がとられるようになりました.

このような「推定」という手法を取る時に，“データ数”のまま分析するよりも“データ数-1”で計算してあげたほうがよりよい推定ができる，ということで“データ数-1”をするようになりました。

本当はもう少し細かな数学的な議論があるのですが，入り込むと帰って来れなくなるのでここまでにしておこうと思います。とりあえず，これからは「不偏分散・不偏標準偏差」が使われることが多い，とだけ覚えておいて下さい。

興味のある方はコチラをご参照ください。

分散が一致することを確認する。

```
var(weight) # 分散
```

```
## [1] 138.2381
```

```
huhens_weight<-sum_hensa_weight2/(length_weight-1)
huhens_weight
```

```
## [1] 138.2381
```

```
sd(weight) # 標準偏差
```

```
## [1] 11.75747
```

```
huhens_hweight<-sqrt(huhens_weight)
huhens_hweight
```

```
## [1] 11.75747
```

## 演習問題 7 :

以下のデータ 3 つ数値の平均値、偏差平方和、分散、標準偏差を求めてください。なお、分散および標準偏差については不偏分散・不偏標準偏差でかまわない。

- 「2, 3, 3, 4」なお，これらのデータは「data1」というオブジェクトに入れて算出すること
- 「1, 1, 1, 9」なお，これらのデータは「data2」というオブジェクトに入れて算出すること
- 「43.6, 45.2, 45.4, 45.8, 47.2, 47.8, 48.2, 48.7, 48.8, 48.9, 49.0, 49.0, 49.4」なお，これらのデータは「data3」というオブジェクトに入れて算出すること

## 今日の課題

### 今日の課題

- 各ページにある演習問題を行う。

- 今日作業した.R ファイルを提出してください.
- 提出方法は案内する.

## おうちでやっておこう

以下の動画を見て作業をしてください.

- R および RStudio のインストールをしておきましょう (いずれも明治大学データ解析論 I の講義資料より).
  - R のインストール : <https://www.youtube.com/watch?v=tJQmdNXkeso>
  - RStudio のインストール : <https://www.youtube.com/watch?v=9AjdZaYOQp4>

## 次回の案内 :

次回の案内 :

- 次回は 2 種類の分散・標準偏差の説明をした後に, 様々なデータの可視化を学びます.
- また, データの扱い方を含めた「実証分析の方法」についても紹介します.

## R でデータを扱う時に注意すべきこと

R でデータを扱う時に注意すべきこと

- 必ず数字／文字は半角で入力する.
- 日本語は使わずにローマ字を使用する.
- コメントアウト (コードではなく, 関係ないメモを入れること) をするときは半角の「#」から始める.
  - メモする内容は全角でもよい.
- ファイル名およびパスには決して全角の文字 (ひらがな, カタカナ, 漢字, 全角スペースなど) を入れてはいけない.
  - 半角英数字だけにする.
- 慌てずに落ち着いて操作すれば, 決して難しくない.
  - 1 つずつ落ち着いて作業することを心がける.
- 「わからない」ことを恐れない
  - 周りの友人に聞いたり, 教員に確認したりしよう.