

確率統計と情報処理・演習 (2008 年度後期)

2008 年 09 月 26 日

日本女子大学理学部数物科学科 今野 良彦

September 25, 2008

今野 良彦

確率統計と情報処理・演習 (2007 年度後期)

この講義の目的と概要

目的

統計学とは、数量的なデータからそのデータの由来する現象に関する情報を科学的に取り出す方法とその理論体系である。また、近年、コンピュータの急速な普及により、あらゆる分野で統計的データ解析の重要性が認識されている。データ解析手法を使いこなすためには、統計学の背景にある科学的手法としての理論を理解することが重要である。このような観点からいくつかのテーマについて学び、演習時にコンピュータによる実習も行う「確率統計と情報処理演習」とすること。

コンピュータとの関連

この講義では、コンピュータ環境 R を用いて実習を行い、統計学の背景となる理論の理解を深める。

1

今野 良彦

確率統計と情報処理・演習 (2007 年度後期)

教科書

垂水共之・飯塚誠也:

R/S-Plus による統計解析入門, 共立出版 (ISBN 4-320-01807-9)

2

今野 良彦

確率統計と情報処理・演習 (2007 年度後期)

09月26日	(1) ガイダンス (2) R の起動と終了 (3) データの入力, 修正, 保存, 読み込み
10月03日	(1) 1次元データの分布 (2) 代表値 — 平均値と中央値・最小値と最大値 (3) ヒストグラム・箱ひげ図 (4) はらつきの尺度 — 範囲・四分位範囲・平均偏差・分散・標準偏差
10月10日	(1) 2変量のデータの分析 (2) 散布図について (3) ファイルからのデータ入力 (3) 回帰直線と最小2乗法について (4) ピアソンの相関係数の性質 (5) 順位相関係数 (6) 多変量データのグラフ表現
10月24日	(1) 予備日 (復習)
10月31日	(1) 確率の定義と性質 (2) 確率関数とその分布 (3) 離散型確率変数と確率関数 (4) 連続型確率変数と確率密度関数 (5) 累積分布関数 (6) 期待値と分散
11月07日	(1) 二項分布の定義とその性質 (2) 関数のグラフ (3) 正規分布と一様分布 (4) χ^2 分布 (5) t 分布 (6) F 分布 (7) 2変量正規分布
11月14日	(1) 予備日 (復習)
11月21日	(1) 標本分布について (2) 有限母集団の標本分布 (3) カイ自乗分布の確率密度関数 (4) 自由度 1 のカイ自乗分布と標準正規分布の関係 (5) カイ自乗分布の再帰性について
11月28日	(1) 一様分布と中心極点定理 (2) さまざまな分布と中心極点定理
12月06日	(1) 母集団と標本 (2) 点推定 — 不偏性・一致性・有効性・最尤法 (3) 正規分布の母平均の点推定 (4) 正規分布の母分散の点推定
12月12日	(1) 信頼度と区間推定 (2) 正規分布の母平均の区間推定 (3) 正規分布の母分散の区間推定
12月19日	(1) 検定の考え方 (2) 正規分布の母平均の検定 (3) 正規分布の母分散の検定
01月09日	(1) 検出力について (2) 復習

3

今野 良彦

確率統計と情報処理・演習 (2007 年度後期)

宿題の提出について

- 提出はメール: mtoukei(at)mp[dot]jwu[dot]ac[dot]jp
- メールの件名 (subject) には、ローマ字の名前と苗字と学籍番号, 締め切り日を書くこと。例:
Mejiro Hanako 201700000 due 2008-10-03
- 解答はテキストファイルに記述し, それをメールに添付すること。ファイル名を
ローマ字の名前-締め切り.txt (例: mejiro-hanako-081003.txt)
とすること。グラフは pdf file で添付し, それぞれの pdf file にはどのようなグラフがあるかをテキストファイルに解説すること。
- Word file では提出しないこと!

4

今野 良彦

確率統計と情報処理・演習 (2007 年度後期)

R について

- S 言語・環境を見本に開発されたフリーソフト (無償で使え, Windows, Mac, Linux)
<http://www.r-project.org/>
- Graphics 機能の充実 (画面または印刷物への出力)
- データを効率的に操作し, 保管する機能
- データの配列機能と列の計算のための演算子のセット
- たくさんのデータ解析の手法
- 条件分岐, ループ, ユーザー定義の再帰的関数 (プログラミングができる。C 言語に似ている)

5

今野 良彦

確率統計と情報処理・演習 (2007 年度後期)

R の基本操作 — データの入力と修正

起動するには

デスクトップの R のアイコンをクリック

データの入力

番号	身長	番号	身長
1	148	9	137
2	160	9	149
3	159	10	160
4	153	11	151
5	151	12	157
6	140	13	157
7	156	14	144

6

今野 良彦

確率統計と情報処理・演習 (2007 年度後期)

データの入力

```
> height<-c(140,150)
> height
[1] 140 150
> height<-c(
+ 140,
+ 150
+ height
[1] 140 150
>
```

オブジェクトの検索

```
> weight<-c(40,50,54)
> weight
[1] 40 50 54
> weihgt
エラー: オブジェクト "weihgt" は存在しません
> ls()
[1] "weight"
>
```

7

R の基本操作 — データの追加・削除・修正

データの追加・削除・修正

```
> h<-c(10,20,30,40,50)
> h
[1] 10 20 30 40 50
> newh<-append(h,c(60,70))
> newh
[1] 10 20 30 40 50 60 70
> h2<-c(newh[2:4],100,110)
> h2
[1] 20 30 40 100 110
```

8

データの保存

メニューの「ファイル」から「ディレクトリの変更」を選び、「Browse」からプロッピーディスクのあるディレクトリを選択する。

```
> sink("result.txt")
> # result.tex という名前のファイルに以下の結果を書き込む
> height
> mean(height)
> max(height)
> sink() # 書き込みを中止する。
```

result.tex の内容

メニューの「ファイル」から「ディレクトリの変更」を選び、「Browse」からプロッピーディスクのあるディレクトリを選択する。

```
[1] 140 150
[1] 145
[1] 150
```

9

データと関数の保存

メニューの「ファイル」から「ディレクトリの変更」を選び、「Browse」から MO のあるディレクトリを選択する。起動から終了までに作成したデータや関数が保存できる。

```
> cc<-c(10,20,39)
> save.image("intro.RData")
>
```

メニューの「ファイル」から「ディレクトリの変更」を選び、「Browse」からプロッピーディスクのあるディレクトリを選択する。

```
[以前にセーブされたワークスペースを復帰します]
> load("intro.RData")
> cc
[1] 10 20 39
>
```

11

演習

以下の実行結果を提出せよ。

- オブジェクト名をつけて、そのファイルに身長データを入力せよ。
- 最後のデータを自分の身長データで置き換えよ。
- 締切り 2008 年 10 月 03 日 (金) 13:00

12