

# 統計-1 (2)

高久雅生

[masao@slis.tsukuba.ac.jp](mailto:masao@slis.tsukuba.ac.jp)

2013年11月20日(水) 1・2時限

# 前回の出席票から (1)

## 【教室設備】

- スライドを映すときは前の電気を消して欲しいです。
- スクリーンが見つらなかった(黒板上の電気がついているせい?)。
- 教室前方の電気を消して頂けるとスライドが見やすいので、消していただけたら...と思います。
- スクリーンが見つらいので、最前列の蛍光灯を消して欲しいです。
  - ✓ 了解しました！以降、気をつけます。
- 教室内の暖房を入れて頂くことは可能ですか
  - ✓ 205教室の暖房を入れることは可能と聞いています。付け忘れていたら、適宜、自分たちで入れるようにしていただけると助かります。
  - ✓ ※どうしても朝一番は温まりづらいと思うので、寒さ対策も忘れずに。

## 【評価】

- 提出課題と期末テストの評価割合はどのようですか？
- 出席票は評価に関わりますか？
  - ✓ 目安として 提出課題: 3割, 期末テスト: 7割, 出席: 1割 とする予定です。
  - ✓ ただし、成績評価は総合的に行います。

# 前回の出席票から (2)

## 【授業資料】

- 統計-1のレジユメをMoodleからテキストを印刷し、それに授業を聴きながらメモしていこうと思っております。授業前に印刷する必要があります。資料をアップする曜日や時刻の目安などあれば教えてください。
  - ✓ できるだけ、前日の夕方にはあげられるようにしたいと思います。しかしながら、経験的に、授業で使う資料の内容を確定するのはその直前になる場合もありますので、正確な目安をお伝えするのは難しいです。ドラフト版でよければ、前日の夕方くらいにはファイルをMoodleにあげるようにはしたいと思います。ご参考に。

## 【教科書、参考書】

- 教科書(ゼロから学ぶ...)とは別に、書籍部につんである新書(統計でウソをつく法)は購入したほうがよいのでしょうか?
  - ✓ はい。こちらも課題図書として、冬休み課題の必読文献とする予定です。なお、購入する必要はありませんので、何らかの手段で読めるようにしていただければそれで結構です。例えば、図情図書館での必読図書として指定しておきましたので、そちらで代替していただいてもかまいません。
- 先日「統計学が最強の学問である」という本を読みました。頻度論とベイズ論のところしか理解できなかったのもので、この授業を受けてからもう一度読み直したいです。
  - ✓ 前日も話したとおり、統計(<データサイエンティスト<ビッグデータ)はここ1・2年のホットなキーワードとなっていますので、関連書籍や雑誌等での特集も多く組まれています。ぜひ関連する内容があれば手にとって見てください。

# 前回の出席票から (3)

## 【授業の進め方】

- 丁寧で分かりやすいが、もう少しテンポよくスライドを進めて欲しい
- スライド数の割にテンポが良くなく、内容が頭に入ってこなく、1, 2時限目でこの授業はととてもキツイと思いました。どうして説明にそんなに時間をかけるのですか。教科書があるので、ただの説明なら講義の必要が無いと思うのですが。こんなにつまらない授業は初めてです。
  - ✓ 申し訳ないです。授業の進度に応じて調整はしたいと思います。ただ、今年の授業についてはやり方を大きく変えることは難しく、できないと思います。
  - ✓ また、テンポについてはおそらく両極端の意見があり、テンポを早めると必要な説明を省略してしまう恐れがあるため、あまり積極的になれません。

# 前回の出席票から (4)

## 【度数分布、ヒストグラム】

- 授業スライドの14枚目、レンジと度数、階級値のところ、 $a_0$ と $a_n$ はRより少し外に定めると言うことでしたが、Rを $n$ 等分するだけではだめなのではないでしょうか？ Rを $n$ 等分して階級を定めるとどのような不都合を生じるのでしょうか？
  - ✓ さほど不都合は生じません。この教科書の著者による流儀とってください。一般的には、Rを $n$ 等分する方式のほうが一般的なようです。このずらす方式では、上端・下端に重なるデータ項目が生じないため、どの階級に属するか容易に見極められる点が簡単になるため、これを採用しているようです。
- 度数分布表を作成する際のデータの集計を効率よく行うコツが知りたいです。プログラミングを使ってできたりするんですか？
  - ✓ Excelのような表計算ソフトウェアや、SPSS・R等の統計用ソフトウェアを使って行う方法もあります。

# 前回の出席票から (5)

## 【階級数・階級幅の求め方】

- $n = 1 + \frac{\log_{10} N}{\log_{10} 2}$  というのは、何か明確な理由があつて、決められているのですか？
  - ✓ はい、一応あります。それは、データが正規分布(後の回で解説します)となることを仮定した上で、2のべき乗からなる個数のデータの場合には、ヒストグラムが対称形となるように定めた、という前提があるようです。実はそれ以外ではあまり意味を成しません。
- 階級数を求める公式が正しい(適正だ)という理由は何か定められているのでしょうか？
  - ✓ 統計学的に、「正しい」「適正な」階級数というモノは存在しません。あくまでも、データの分布が分かりやすく合理的に見えればよいということが目標になります。この目標を満たすためにはデータの性質を考慮する必要があります。
- スタージェスの公式や鈴木公式は参考までにということですが、覚える必要はありますか？
  - ✓ 特に覚える必要はありません。
- 階級数を決めるときに、経験則を用いる場合、スタージェスや鈴木公式は計算が難しいので、おおよその見当を利用するようにしても大丈夫でしょうか？
  - ✓ はい、おおよその見当を用いることでもかまわないと思います。
- スタージェスの公式と鈴木公式は併用すればよいのでしょうか。それとも、どちらか片方を使えばよいのでしょうか。
  - ✓ 併用する必要はありません。

# 前回の出席票から (6)

- スタージェスの公式では常用対数が使われていますが、試験時に対数の値は与えられるのでしょうか？ それとも目安を暗記して解かなければならないのでしょうか？
  - ✓ 公式も目安も暗記の必要はありません。もし必要であれば対数の値は提示します。
- 階級数として  $\log_2 N$  は整数になることが少ないと思うのですが、四捨五入してよいのでしょうか？
  - ✓ 確かに、非整数の階級数というものは存在し得ないので、整数にする必要があります。ただし、四捨五入は使わず、当該値を超えた最小の整数 (Ceiling function) によって階級数を定めることが多いようです。
- 階級幅は、人によって違うということでもいいんですか？
  - ✓ はい、人により異なる階級幅が定義されることは自然な結果だと思います。
- 階級の分け方が難しいと思いました。統計の作成者により、統計は違ってくると思いますが、提出課題や試験において、ある程度の解答の幅は認められるのでしょうか？
  - ✓ はい、その解答の幅があることは全く問題ありません。ただし、その解答を導いた理由を合理的に説明できるようにしてください。

# 前回の出席票から (7)

## 【その他】

- 他学類として教職としてこの授業を受けることは可能ですか？
  - ✓ はい、本科目は特に受講の制限は設けていません。
  - ✓ 知識情報・図書館学類における、数学の教科に関する科目としても指定されていますが、他学類からの受講であっても本学類の数学科目を受講してこれを申請することは可能です。ただし、実際の教員免許の申請は所属学類の支援室が対応しますので、あらかじめ教職担当に確認、相談するようにしてください。
- 仮説検定という考え方は、統計学の入門的な概念なんですか？
  - ✓ その通りです。「仮説検定」は、統計学におけるもっとも基本的な概念の一つです。場合によっては、科学における基本的な概念の一つと言っても良いかもしれません。
- 第1回課題の『受講生60名』とは、「受講生(合格者)全60名」か「受講生(合格者)のうち60名」か？
  - ✓ 後者です。課題例のために抜粋したものです。



# 本日のお品書き

- 代表値
  - ✓ 平均
  - ✓ メディアン
  - ✓ モード
- 散布度
  - ✓ 分散
  - ✓ 標準偏差
- 相関と回帰

# 第2回課題

A) 以下の数列は2013年度某大学の某授業科目における成績得点の抜粋である。

1. 平均、メディアン、モードを求めよ。
2. 分散、標準偏差を求めよ。

77 78 80 83 78 80 85 78 76 78 62 80
-------------------------------------

## 第2回課題 (2)

- B) 以下の度数分布表は2013年度某大学の某授業科目における成績得点の度数分布である。
1. 平均、メディアン、モードを求めよ。
  2. 分散、標準偏差を求めよ。

階級値	25	35	45	55	65	75	85	95
度数	1	2	0	2	9	81	24	4

# 第2回課題(提出方法)

- 課題はA4用紙(1枚)にて提出すること。
  - ✓ なお、2ページにわたる場合は裏面に記載のこと。
- 科目名(統計-1)、課題名(第2回課題)、提出年月日、学籍番号、所属、氏名を提出用紙の一番上に必ず記入すること。
- 提出×切:2013年11月27日(水)
  - ✓ 授業時間中に提出を求めます。
  - ✓ やむをえず欠席等する場合は、7D 208研究室前に提出場所を用意するので、そちらに提出すること。

# 代表値、散布度、相関



教示用スライドへ

# 本日のまとめ

- 代表値
  - 平均
  - メディアン
  - モード
- 散布度
  - 分散
  - 標準偏差
- (相関と回帰)
  - 相関係数
  - 回帰