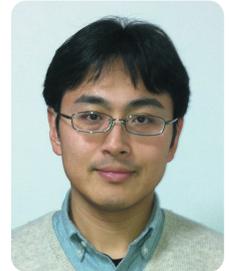


人材育成  
のための  
授業紹介

統計学

# JMOOCプラットフォームgaccoの講座 「統計学Ⅰ：データ分析の基礎」について



中央大学  
理工学部准教授 酒折 文武

## 1. はじめに

情報通信技術の高度化・深化に伴って、社会で得られるデータは爆発的に増大し、いわゆるビッグデータ時代が到来しました。そのような中で、データに基づいた客観的な判断と意思決定は、ビジネスや産業界のみならず、現代社会に生きる市民に広く必要とされています。

そこで日本統計学会では、データを正しく解釈し科学的な問題解決を行うための統計学のオンライン講座を開設することになりました。総務省統計局およびNTTデータグループの協力のもと、JMOOCのプラットフォームの一つgaccoで講座を提供し、筆者らが講師を務めました。JMOOCやgaccoについては、本誌バックナンバーの特集<sup>1)</sup>をご覧ください。本稿では講座の授業内容やその反応などを紹介します。なお、本稿の内容は筆者単独の見解に基づいており、gaccoや日本統計学会あるいは担当した他の講師の方々に文責はないことを付記しておきます。

## 2. 講座の概要

講義は2014年11月12日から5週間かけて行われ、講師として4名の大学教員に加えて1名の総務省統計局職員が担当しました。筆者は第3週の担当です。各週のテーマと内容は表1の通りです。一般に「記述統計」と呼ばれる内容を中心として、随所に実用上重要なトピックを含めました。その結果、各週の内容は通常の大学の授業1コマでは収まりきれないほど幅広く充実した内容となっています。

各週は10分×8本程度の動画による講義と、確認テスト（第1～4週）、最終テスト（第5週）

から構成されます。そして、受講者同士、あるいは講師やTAも交えた議論の場としての掲示板「ディスカッション」が設置されています。また、講座の副読本（別途申込、有料）も作成しました。そして、希望者を対象とした反転学習コース(有料)も設置しました。以下で順に紹介していきます。

なお本講座は、提供している日本統計学会をはじめとする統計関連の学会が策定している「統計学分野の教育課程編成上の参照基準」の体系に沿うものです。また、この基準に基づいて認定している「統計検定」の資格検定の内容にも対応しています。

表1 講座の内容

週	テーマ	内容
1	統計学への誘い	全体の概観。データに基づいた意思決定の重要性、統計学の歴史、データ分析の流れ、データの種類
2	統計グラフと質的データの要約	質的データをグラフにまとめる、クロス集計、連関関係・因果関係
3	量的データの要約	量的データをグラフにまとめる、代表値や散布度、偏差値、ローレンツ曲線
4	相関と時系列	量的な2変数の関係、単回帰分析、時系列データの指数・変化率・成長率
5	公的統計の活用とまとめ	公的統計や統計調査の仕組み、政府統計ポータルサイトの活用、全体の総括

## 3. 動画による講義と確認テスト

オンライン講座では、1) 教室での通常の講義（板書あるいはスライド）を映したものの、2) ス

ライドと講師が同時に映し出されるもの、3) 画面にはスライドだけが映って声がそれに重ねられているもの、4) カーン・アカデミーのように画面上に次々と式や文字が書き加えられていくもの、など様々なタイプの講義があります。gaccoの講座では、このうち2)の形式がとられます。大学の通常の授業で、スライドだけでなく黒板・タブレットへの板書やオンライン上のアプレット等を随時活用して講義していた筆者は、少々やりにくさを感じつつも、なるべく分かりやすいスライドとなるよう配慮して準備を進めました。図1は、筆者による実際の講義動画の一例です。



図1 動画による講義の例

本講座では、各週の確認テスト10問×4週と、最終テスト30問(各2点)の合計100点で評価を行い、60点以上を合格としました(実際には出題ミス等で点数調整がありました)。

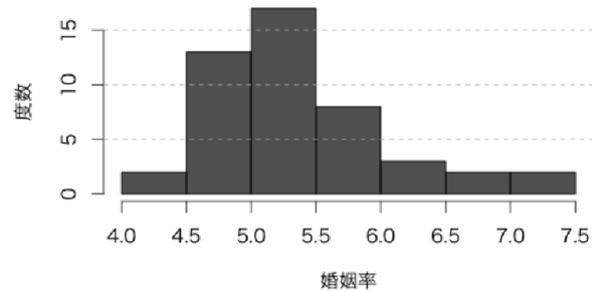
各週の確認テストおよび最終テストは、5択のクイズ形式です。統計リテラシーすなわち統計学の基本的な用語や概念の理解と、統計的推論すなわち統計情報を読み取って解釈・推論する力、の両方を問う問題を出題しました。図2は実際の確認テストの例です。この問題はヒストグラムから読み取れることを確認することが目的です。

大学の講義とは異なり対面ではないことから出題意図が伝わりにくく、受講者が戸惑う問題もあったようです。また、準備期間が非常に短かったため、少数ではありますが誤植や出題ミスがあり、受講者から指摘を受けたりもしました。

なお、gaccoはOpen edXという世界標準のMOOCシステムを採用しており、その中に記述式の課題を受講者が相互採点する機能があります。今回の講座ではこの機能を活用しませんでした。今後は検討してもよいかもしれないと個人的には感じています。

問題8 (1点満点)

【問題8】日本の47都道府県における、ある年の婚姻率のデータを調べたところ、次のようなヒストグラムを得た。



ただし、婚姻率とは、人口1,000人当たりの婚姻件数として定義されている。このデータから読み取れる婚姻率に対する記述として適切でないものを、次のa.~e.のうちから一つ選びなさい。

- a. 婚姻率が6.5以上の都道府県が4つある
- b. 婚姻率の中央値は、5.0以上5.5未満の階級にある
- c. 婚姻率の平均値は、5.0以上5.5未満の階級にある
- d. 最も度数が多いのは、5.0以上5.5未満の階級である
- e. 婚姻率が5.5以上の都道府県の割合は30%以上である

図2 確認テストの例

#### 4. ディスカッション

ディスカッションと呼ばれるスレッド型掲示板は、受講者同士での議論や、講師・運営側への質問などに活用されます。講座開始直後から連日、様々なスレッドが立ち上げられ、非常に多くのコメントで賑わいました。次ページの図3はその一例です。多くは講義内容や確認テストについての疑問でしたが、それらの大部分は受講者同士のやりとりで解決されていました。時には非常に明快な回答がなされ、こちらが感心させられることもありました。その他、講義内容への感想や要望、誤植の指摘、ミートアップ(自主勉強会)の企画など、実に多岐にわたる話題で情報交換がなされていました。通常大学の講義と比較して、受講者の皆さんの幅広いバックグラウンドや強い熱意を改めて感じました。

最終的なスレッド数は229にも上りました。講座の期間は1日何回もディスカッションをチェックする毎日でしたが、それでもすべての書き込みをチェックするのは大変で、いろいろな意味で答えに窮する(漠然とした質問、どこまで教えてよいかの判断等)ものも少なくありませんでした。正直なところ、このディスカッションへの対応は想像以上の負担がかかりました。しかし、教

育効果を上げるためには不可欠なのは間違いありません。



図3 ディスカッションの例

## 5. 副読本「スタディノート」

通常の大学での講義と比べてのオンライン講座のメリットは、わかりにくいところを何度も繰り返して学ぶことができることでしょうか。一方、デメリットとして、動画の視聴のみでどうしても受動的に受講しがちになってしまうことが挙げられます。これを解決するために、副読本「スタディノート」を作成しました。

スタディノートでは、講座のすべてのスライドとともに、各スライドの要点をまとめた穴埋め式の説明を加えました(図4)。講義を聴きながら空白部分のキーワードを埋めたり、内容のメモを欄外にとったりと、能動的な学習ができるように工夫しました。さらに、確認テストの模擬問題となる練習問題30問とその解説、講義で説明した分析を実際にExcelや統計ソフトRで実行するための操作法、統計用語集、反転学習のための資料などを掲載しました。多くの受講生に活用してもら

うことができ、評判もよかったです。

## 6. 反転学習コース

統計的な問題解決力を身につけるには、統計リテラシー・統計的推論に加えて、統計的思考すなわち問題の認識からデータ収集・分析・解釈に至るまでの一連の過程を理解し実践する力の育成が欠かせません。オンライン講座では、この統計的思考を育成するのが最も難しいでしょう。

そこで、gaccoにおける他のいくつかの講座と同様に、本講座でも「反転学習コース」を実施しました。反転学習コースは、通常のオンライン講義に加えてアクティブラーニングの対面授業を行う

### 第3回 分布を読みとる

#### さまざまな分布の形

典型的な分布の形について紹介  
分布の形を把握することは正しい分析を行うために非常に重要

単峰型の分布

多峰型の分布

高原型の分布

離島のある分布

右に歪んだ分布  
(右の裾が長い分布)

左に歪んだ分布  
(左の裾が長い分布)

#### 単峰か多峰か

単峰型の分布

2013年の日本プロ野球セバ両リーグの規定打席別進打者の打率

データの集中する山の部分が1つ

とくに左のような左右対称な形をベル型の分布という一般的によく現れる形

多峰型の分布

データの集中する部分(峰)が複数ある形状

オールド・フェイスフル・ガイザー(有名な間欠泉)が次に噴き出すまでの時間

- ・ヒストグラムを作ると、図のようにさまざまな分布の形が現れる。
- ・今回は典型的な分布の形を紹介する。
- ・分布の形を把握することは、数値で要約するなど今後の分析においても非常に重要
- ・打率のデータ(上のグラフ)のように、山が一つの分布を   型の分布という。
- ・単峰であり、ほぼ左右対称の形の場合、  型の分布という。一般的によく現れる形。
- ・間欠泉のデータ(下のグラフ)のように、山が複数ある分布を   型の分布という

図4 スタディノート

ものです。なお、筆者は対面授業の講師を担当していませんので、以下の内容の多くは担当講師から伝え聞いたものです。

対面授業は、東京・大阪・福島の3カ所で2014年12月20日に行いました。申込者は合わせて180名にも上り、男女比7対3、20代前～70代までと幅広い年齢層（うち9割は25～60歳の社会人）の方々が参加されました。



写真1 反転授業の風景

対面授業の全体的な流れは表2の通りです。まず、受付の済んだ参加者を順次6名のグループとなるように、島の形で座ってもらいました。次に、オンライン講義の全体の概要の復習と、掲示板で質問が多かった四つの項目（統計グラフ：誤用と悪用、シンプソンのパラドックス、擬似相関と偏相関係数、データの変換）についてスライドで簡単に説明しました。続いて各グループでまとめ役を互選し、これらの問題点についてどのように考えるかを30分ほど議論してもらいました。その後、各グループの議論の内容をそれぞれ発表してもらいました。社会人が多く、実社会の経験にもとづいて解釈する姿勢が見られ、大学の学生の反応とは少し違う面があったため、講師としても参考になりました。それから各項目について解説資料を追加で配布して、再度詳しく説明しました。

さて、参加者には事前に、任意でデータ分析例の提出をお願いしていました。その中から、アメリカと日本の株価の変動に関する優れた分析例を紹介しました。そして最後に、配布資料に基づいてRによる統計解析の実習を行いました。計画では、最初に通り講師が画面で実行例を示し、その後各グループで実習する形を考えていましたが、参加者からの要望で、各実行例について画面で示しつつ実習も行う形となりました。参加者は概ね熱心でしたが、レベルに差があり、10名に1名程度配置したTAも対応に苦慮していました。今後に向けて実行方法に多少改善が必要かもしれ

ません。

3時間半という限られた時間ではありましたが、データ分析課題や分析実習を通じてオンライン講座の欠点を補うとともに、活気ある活動や議論により参加者の理解を深めることができました。

表2 対面授業の流れ

1	受付とグループ分け
2	オンライン講義の復習と問題点の指摘
3	グループディスカッションとグループ発表
4	問題点の解説
5	受講者によるすぐれたデータ分析例の紹介
6	Rによる統計解析の実習

## 7. おわりに

本講座は1万3千名以上の登録があり、合格者はその15%以上にも上りました。統計学への興味の高さが伺えます。これを受けて、本講座の再放送や続編の作成などの動きも始まりました。また、本講座をベースとして、さらにビジネスとしての活用を修得するコース「反転型研修」も開始されました。

こうしたニーズはこれから益々高まるでしょう。今回の経験とこれまで述べてきたような課題を踏まえて、さらにより講座を提供できるように（筆者個人としても学会としても）努力していきたいと思っています。

先に述べたように、統計的思考の涵養には、受動的な講義の聴講だけでなく、実際にデータ収集や分析を自主的に行うことが重要です。限られた授業時間を有効に活用するためには、大学での正課の科目においても、オンライン講義や反転授業の活用が重要となってくるでしょう。今後の正課科目において、本講座自身あるいは本講座を通して得られた知見の活用を検討するとともに、講義内容と学習成果の対応についても検討評価していきたいと思っています。

## 参考文献および関連URL

- [1] 福原美三: JMOOCを理解するために. 大学教育と情報, 2014年度No.3, pp.2-10. 2014.  
[http://www.juce.jp/LINK/journal/1501/pdf/02\\_01.pdf](http://www.juce.jp/LINK/journal/1501/pdf/02_01.pdf)