

《研究ノート》

コロナ禍における統計教育 —CSI のオンデマンドでの統計教育を踏まえた改善—

山口 誠一
山口 和範
門田 実

【要旨】 世界的な新型コロナウイルスの感染拡大に伴い、2020年度の大学教育が遠隔授業を中心としたものになった。対面での講義の重要性はあるとしても、データ分析のスキルの習得においては、学習者が各自のペースに合わせて、繰り返し視聴し確認しながら学びを進めることができる点で、オンデマンド形式での学びにも大きなメリットがあるといえる。人文社会科学系学部での統計学の基礎のクラスは、クラスサイズが大きくなることが多く、また学習者の数学やコンピュータスキル等の基礎力のばらつきは大きい。一方、統計的思考力の育成を目的とする学習においては、統計などのエビデンスを用いたコミュニケーションの機会の提供が重要になるが、オンデマンド形式では、このような機会を提供することが難しくなる。本報告では、2020年度の立教大学における統計学に関するオンデマンドを主とした遠隔授業の事例を取り上げ、そこで実施されたコンテンツの紹介と、学習者の事前と事後の調査からみる教育効果についての確認、さらには残された課題を示す。

キーワード：遠隔授業，オンデマンド授業，データサイエンス教育

I はじめに

2020年度は、新型コロナウイルス感染拡大に伴う措置として、多くの大学で遠隔授業が展開されることとなった。立教大学でも、春学期についてはすべての科目が遠隔授業として開講され、秋学期もほとんどの科目が遠隔授業として展開された。社会情報教育研究センター（CSI）では、2010年の開設当初より全学部の学生向けの社会調査や統計学の科目をオンデマンド授業として提供してきた実績もあり、従来から提供してきた科目については、対面でのスクーリングが実施できなかったことを除けば、これまでとほぼ同様の教育の提供ができた。統計教育においては、その基礎となる数学や情報科学に関する学習履歴のばらつきに対応することからも、e-learningを含むオンデマンドでの教育の利点が指摘されてきた（渡辺ほか, 2000）。

データ分析のスキルの習得においては、学習者が各自のペースに合わせて、繰り返し視聴し確認しながら学びを進めることができる点が、大きなメリットといえる。一般に、人文社会科学系学部での統計学の基礎のクラスは、クラスサイズが大きくなることが多く、また学習者の数学等の基礎力のばらつきは大きくなる。また、Excelや統計分析専門のソフトウェアのSPSSやRを用いるようなパソコン演習を取り入れた際には、パソコン操作の習熟度に加え、使用するソフトウェアの習熟度にも大きなひらきがあることが多い。一方、統計的思考力の育成を目的とする学習においては、統計などのエビデンスを用いたコミュニケーションの機会の提供が重要になるが、オンデマンド形式では、このような機会を提供するこ

とが難しくなる。なお、遠隔授業形式でも、Zoomなどを活用したリアルタイムで集合しての授業では、対面授業以上に密なグループディスカッションが可能となる。コロナ禍での遠隔授業の際に用いられるようになったZoomなどの遠隔授業ツールには、少人数でのグループワークを行うための機能が備えられているものがほとんどである。Zoomにおけるブレイクアウトセッション機能がその代表的なものである。比較的クラスサイズの大きな授業において、少人数のグループに分かれてのディスカッションを対面授業で行うためには、机や椅子が可動式であることが求められるなどの制限が加わることがあるが、遠隔授業においてはその点の制約がない。

講義の内容や目的においては、対面授業であることが重要視されるが、上記のような観点で考えると、統計学やデータサイエンス教育においては、必ずしも対面授業でなければならぬという制約はないといえる。

そこで、今回の報告では、コロナ禍における新たな遠隔授業の実践例を取り上げ、その効果を確認するとともに、今後の改善に向けた課題を明らかにする。

II オンデマンド形式の教育実践

本章では、立教大学大学院科目の一つである「統計学1」（2020年度秋学期1開講科目）の受講生に対して行った遠隔授業実践の事例を紹介する。受講生は社会人大学院生が主であり、授業の実施形式は、各受講生が都合のよい時間に動画を視聴できる、いわゆるオンデマンド形式で行った。動画視聴を利用した大学の授業には様々な方法があり、例えば自宅等で動画を視聴してから対面授業ではグループワーク、応用問題や演習等を行う反転授業（花木ほか、2015）やMOOCの利用（岩下ほか、2015）等があるが、「統計学1」では、動画の視聴後に一定期間を設けて選択回答式の問題からなる課題を提出してもらう方法をとった。当該科目は統計学の基礎を統計処理ソフトの利用とともに習得できるような内容となっている。筆者はデータサイエンスに関する教材を将来的に開発することを念頭においており、どのような統計教育法がそれに活用できるのかに関心があるため、2020年4月に発表された数理・データサイエンス・AI（リテラシーレベル）モデルカリキュラムの中の“基礎”にあたるデータリテラシーの学修目標に関する項目を含む授業アンケートを実施した。本章では、当該科目を通じて得た、オンデマンド形式での統計科目の授業運営に関する知見等を紹介するとともに、受講生に対して行ったアンケートの結果についても考察する。

1. 「統計学1」について

2012年以降に高校数学Iで「データの分析」が必修化されるなど、初等中等教育において統計に関する内容の充実化が行われたが、当該科目の受講生は、それらの統計教育を受けてはいないと考えられる社会人が主である。統計学に関する基礎的な内容および記述統計の基礎を固め、統計的推定・検定などの推測統計の基礎およびSPSSやEXCEL等を用いたデータ分析スキルを身につけることを授業目標としている。遠隔授業の実施形式は、各受講生が都合のよい時間に動画を視聴できる、いわゆるオンデマンド形式で行ったが、時間割の上では週に2コマ（1コマ100分を続けて2コマ）が割り当てられており、この時間

の開始に合わせて Google Drive で共有する動画の URL と課題を LMS (Blackboard) 上で公開した。LMS を通じて受講生に各種の案内を行い、質問や学生間の交流の機会の確保等は LMS の掲示板や大学のメールを利用した。コロナ禍において大学に入校できないことで、受講生が SPSS を利用できなくなることが懸念されたが、学外から VPN 接続を用いて利用できる対応がとられた。当該科目で扱った主な内容は、社会における統計 (第 1 週)、記述統計と推測統計の概要 (第 2 週)、1 変数データの要約 (第 3 週)、2 変数データの要約 (第 4 週)、推定 (第 5 週)、検定の考え方 (第 6 週)、平均値の差の検定 (第 7 週) である。第 2 週には、CSI で過去に作成した「SPSS 統計解析 (Basic コース)」という SPSS の基本操作を解説するオンデマンド教材を案内した。GAISE レポート (GAISE, 2005) でも推奨されているように、リアルデータを用いることを重視し、当該科目では、e-Stat や RESAS の利用および東京大学の SSJDA から JGSS-2010 という社会調査データの提供を受けるなどして、リアルデータを分析の演習等に利用した (JGSS2010)。各週必ず課題 (単一選択回答形式の問題) を課し、授業動画公開日から 4 日間程度の提出期間を設け、特に 3 週目以降は JGSS-2010 データを SPSS を用いて分析する課題を出題した。これは、筆者が過去に既に立教大学で開講されていたオンデマンド科目を複数担当し、動画を何時でも視聴でき、課題の提出期限を長くしすぎると、視聴すべき動画をため込む受講生が一定の割合で存在するという経験からそうしている。

2. 数理・データサイエンス・AI (リテラシーレベル) モデルカリキュラム

ここでは、後述するアンケートにおけるデータリテラシーに関する項目作成のために参考にした数理・データサイエンス・AI (リテラシーレベル) モデルカリキュラムを紹介する。統合イノベーション戦略推進会議による AI 戦略 2019 において、AI 時代に対応した人材の育成・確保は、初等中等高等教育だけに限らずリカレント教育や生涯学習をも含めた長期的課題とされ、特に「数理・データサイエンス・AI」の知識・技能と人文社会芸術系の教養をもとに、新しい社会の在り方やサービス等をデザインする能力が重要であり、これまでの教育方法の抜本的な改善および実社会の課題解決的な学習を教科横断的に行うことが不可欠とされている (AI 戦略 2019)。さらに、2025 年に実現することを念頭にした教育目標の中には、「文理を問わず、全ての大学・高専生 (約 50 万人卒/年) が、課程にて初級レベルの数理・データサイエンス・AI を習得」との記述があり、その教材および教育方法の開発は喫緊の課題といえるであろう。数理・データサイエンス教育強化拠点コンソーシアムは、この AI 戦略 2019 の提言等を踏まえ、分野を問わず、全ての大学・高専生 (約 50 万人卒/年) を対象にしたリテラシーレベルの教育の基本的考え方、学修目標・スキルセット、教育方法をまとめた「数理・データサイエンス・AI (リテラシーレベル) モデルカリキュラム」(以下、モデルカリキュラムと略記)を 2020 年 4 月に発表した (モデルカリキュラム, 2020)。このモデルカリキュラムは“導入”、“基礎”、“心得”、“選択”からなり、これらの概要を述べると、“導入”はデータサイエンスに関する社会の変化やデータ・AI の活用領域・技術を知ること、“基礎”はデータを読む・説明する・扱うというようなデータリテラシーに関すること、“心得”はデータ・AI 利活用における倫理やリスクに関すること、そしてアルゴリズム基礎や画像解析、プログラミング基礎など多様な項目から選ぶことができるのが“選択”となっている。これらを各大学・高専の状況に応じて適切に選択・抽出できるように柔軟性

があるカリキュラムになっている。ここでは“基礎”にあたるデータリテラシーをとりあげることとし、図表1にモデルカリキュラムとデータリテラシー<スキルセット>を示す。

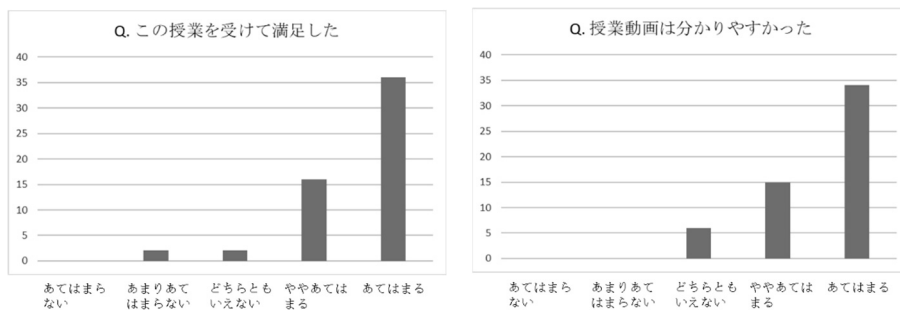
導入	1. 社会におけるデータ・AI活用	
	1-1. 社会で起きている変化	1-2. 社会で活用されているデータ
	1-3. データ・AIの活用領域	1-4. データ・AI活用のための技術
	1-5. データ・AI活用の現場	1-6. データ・AI活用の最新動向
基礎	2. データリテラシー	
	2-1. データを読む	2-2. データを説明する
	2-3. データを扱う	
心得	3. データ・AI活用における留意事項	
	3-1. データ・AIを扱う上での留意事項	3-2. データを守る上での留意事項
選択	4. オプション	
	4-1. 統計および数理基礎	4-2. アルゴリズム基礎
	4-3. データ構造とプログラミング基礎	4-4. 時系列データ解析
	4-5. テキスト解析	4-6. 画像解析
	4-7. データハンドリング	4-8. データ活用実践（教師あり学習）
	4-9. データ活用実践（教師なし学習）	

2.データリテラシー	キーワード（知識・スキル）
2-1. データを読む	<ul style="list-style-type: none"> データの種類（量的変数、質的変数） データの分布（ヒストグラム）と代表値（平均値、中央値、最頻値） 代表値の性質の違い（実社会では平均値＝最頻値でないことが多い） データのばらつき（分散、標準偏差、偏差値） 観測データに含まれる誤差の扱い 打ち切りや脱落を含むデータ、層別の必要なデータ 相関と因果（相関係数、擬似相関、交絡） 母集団と標本抽出（国勢調査、アンケート調査、全数調査、単純無作為抽出、層別抽出、多段抽出） クロス集計表、分割表、相関係数行列、散布図行列 統計情報の正しい理解（誇張表現に惑わされない）
2-2. データを説明する	<ul style="list-style-type: none"> データ表現（棒グラフ、折線グラフ、散布図、ヒートマップ） データの図表表現（チャート化） データの比較（条件をそろえた比較、処理の前後での比較、A/Bテスト） 不適切なグラフ表現（チャートジャンク、不必要な視覚的要素） 優れた可視化事例の紹介（可視化することによって新たな気づきがあった事例など）
2-3. データを扱う	<ul style="list-style-type: none"> データの集計（和、平均） データの並び替え、ランキング データ解析ツール（スプレッドシート） 表形式のデータ（csv）

図表1：数理・データサイエンス・AI（リテラシーレベル）モデルカリキュラム（上）とデータリテラシー<スキルセット>（下）

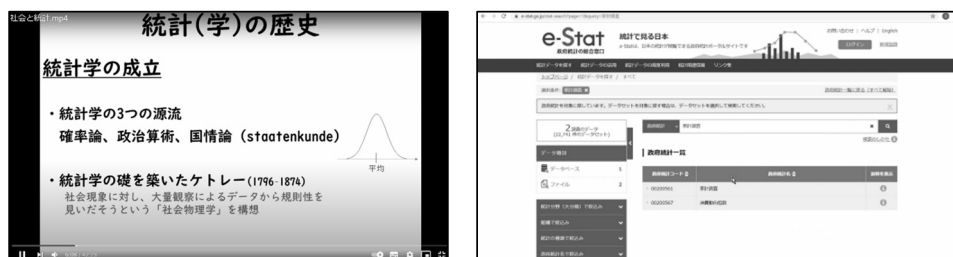
3. 受講生に対して行ったアンケートの内容と結果および考察

「統計学1」の受講生63名に対し、第1週目の授業の直前に、初回の授業時において、まず事前アンケート（pre）に回答することを求める案内をLMSを通じて行った。また、事後アンケート（post）は全ての週の授業が終了した後にLMSを通じて回答を依頼した。LMS上で、事前アンケートにおいて、アンケートの結果を個人が特定されない形で公表することに同意する回答欄を設け、アンケートの回答内容・未回答等で不利益を被らないこと、および同意は撤回できることを伝え、授業終了後に一定の期間を設け、同意撤回の機会を確保した。まずは事後アンケートにおける、授業満足度と授業動画の分かりやすさに関する質問の回答結果を図表2に示す。



図表 2 : 授業満足度 (左, N=56) と、授業動画の分かりやすさ (右, N=55)

アンケートでは、学習内容に関して良かった点と改善が必要だと思う点を自由記述の回答形式でそれぞれ質問したところ、良かった点は「e-Stat や RESAS を活用して情報を収集・分析する方法も学べた」、「論文を輪読する際、分析の結果が提示された時、結果を理解することができ、論文の内容を適切に理解することができました」という記述などがあり、データを正しく読んで活用する力の向上が伺えた。改善が必要だと思う点では「実務に関係しそうな応用例をもっと知りたい」というような授業内容が受講前に想定していた内容と異なっていたという記述が 2 件あった。ほとんどの動画ではナレーション原稿を用意し、PowerPoint の画面録画機能を利用して PC で作成した動画を受講生に視聴してもらった (図表 3)。EXCEL や SPSS の演習も同様の方法で動画にし、可能な限り授業動画の質を高めるよう努めた。



図表 3 : 授業動画の一部 (講義 (左) と e-Stat を利用した演習 (右))

次に、受講前後におけるデータリテラシーの変化を見るため、図表 4 の様にモデルカリキュラムのデータリテラシーにおける各学修目標に対して質問文を作成し、5 件法で質問した (1:あてはまらない, 2:あまりあてはまらない, 3:どちらともいえない, 4:ややあてはまる, 5:あてはまる)。ここでは、この 8 項目全てに回答した 51 名に対し、この質問で得た変数を量的変数とみなして対応のある t 検定および推定の結果を図表 5 に、回答分布を図表 6 に示した。どの項目も概ね受講後に良くなっている。事後アンケートの方の回答分布を見ると、“あてはまる”の回答者の割合にはまだ伸び代が十分ある。今回は 4 月に発表されたカリキュラムのデータリテラシーに関する学修目標をほとんどそのままの形式で質問文に変えて試行的に用いたので、“あてはまる”と自信を持って回答し難い質問であったかもしれないこと

もあり、将来的にはルーブリックの作成等を検討する必要があると思われる。また、Q5が比較的事後に十分に上方にシフトせず、postにおいて最も平均値が低く、回答分布のばらつきが大きかった。これは、可視化を行うまでは授業や演習、課題等で行ったが、他者に説明するという機会が授業内ではほぼ無かったことが要因かもしれない。グループワーク等でデータの説明を行い、また、他者の意見を聞き議論することは、データを多面的にとらえる貴重な経験となるであろうが、対面形式の授業では容易に行えたグループワーク等がコロナ禍では難しくなってしまった。グループワーク等を通じて自身の分析結果等を他者に伝える機会をどのような形で設けるかを検討する必要があると思われる。

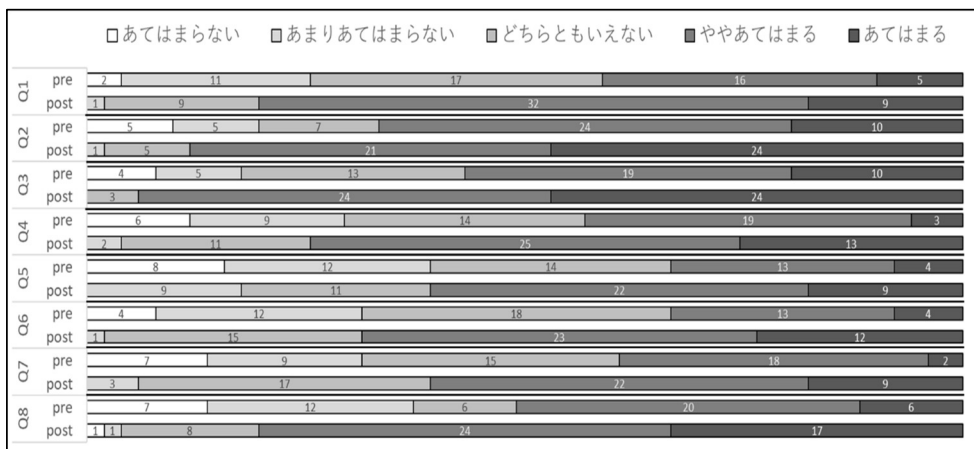
質問文	モデルカリキュラムの学修目標（データリテラシー）
Q1.データの特徴を読み解き、起きている事象の背景や意味合いを理解できる	データの特徴を読み解き、起きている事象の背景や意味合いを理解できる
Q2.データを読み解く上で、ドメイン知識が重要であることを理解している（ドメイン知識：特定の領域のデータを読み解く場合、その領域知識のこと。例えば、菓子業界のデータを読み解いているなら菓子業界の知識など）	データを読み解く上で、ドメイン知識が重要であることを理解する
Q3.データの発生現場を確認することの重要性を理解している	データの発生現場を確認することの重要性を理解する
Q4.データの比較対象を正しく設定し、数字を比べることができる	データの比較対象を正しく設定し、数字を比べることができる
Q5.適切な可視化手法を選択し、他者にデータを説明できる	適切な可視化手法を選択し、他者にデータを説明できる
Q6.不適切に作成されたグラフと数字に騙されない	不適切に作成されたグラフ/数字に騙されない
Q7.文献や現象を読み解き、それらの関係を分析・考察し表現することができる	文献や現象を読み解き、それらの関係を分析・考察し表現することができる
Q8.エクセルまたはスプレッドシート等を使って、小規模データ（数百件～数千件レベル）を集計・加工できる	スプレッドシート等を使って、小規模データ（数百件～数千件レベル）を集計・加工できる

図表4：質問文とモデルカリキュラムのデータリテラシーの学修目標との対応

	平均値(不偏標準偏差)		t値	差の95%信頼区間	
	pre	post		下限	上限
Q1	3.22 (1.026)	3.96 (0.662)	5.234**	0.459	1.031
Q2	3.57 (1.204)	4.33 (0.739)	4.396**	0.451	1.114
Q3	3.51 (1.155)	4.41 (0.606)	5.671**	0.582	1.221
Q4	3.08 (1.129)	3.96 (0.799)	6.103**	0.592	1.173
Q5	2.86 (1.200)	3.61 (0.981)	4.791**	0.433	1.057
Q6	3.02 (1.068)	3.90 (0.781)	5.204**	0.542	1.223
Q7	2.98 (1.122)	3.73 (0.827)	4.573**	0.418	1.072
Q8	3.12 (1.291)	4.08 (0.868)	5.496**	0.610	1.312
Total	25.35 (7.011)	31.98 (4.680)	7.181**	4.774	8.481

** p < 0.01

図表5：対応のあるt検定および差の信頼区間（N=51）



図表 6 : Q1~Q8 の pre-post の回答分布 (N=51)

アンケートでは、学習形態に関して良かった点と改善が必要だと思う点も自由記述の回答形式でそれぞれ質問した。良かった点については、この質問に回答した 51 名中 31 名が何度でも動画を視聴できる点に触れており、昨年度までの一度だけの受講という形式と比べたこの変化は学生に大きなメリットをもたらしたと考えられる。これはオンデマンド形式だけのメリットではなく、リアルタイムの遠隔授業を録画して受講生に公開するというような場合も同様であろう。リアルタイムの授業の録画とオンデマンド形式のような事前に作成する動画との大きな違いの一つは、オンデマンド形式の方では、言い忘れ、言い間違い、書き間違い等のミスを極力減らすことが可能であり、動画の視聴のしやすさや分かりやすさの向上を事前に行うことができることである。改善が必要だと思う点としては、質問がしにくいという意見が 5 件あり、質問のしやすい環境づくりを検討する必要がある。また、自由記述から読み取れたこととして、好きな時間に動画を視聴してもよいオンデマンド形式の授業は、各受講生の学習スケジュールの調整機能を果たしうるということである。コロナ禍において始まった遠隔授業では、昨年度までと比べて課題の量が大幅に増えたと言われているが、オンデマンド形式の授業が学生の負担を緩和した可能性が考えられる。そして、コロナ禍で世界中で作成されたオンデマンド形式の授業のための大量の動画は、高等教育における貴重な財産であることも指摘しておきたい。優れた動画を共有するなどし、各大学が動画の視聴で行うことができる教育が何かを見極め、高等教育がより良くなる方向に有効に活用すべきであろう。

III 結語

2020 年度の大学教育は、コロナ禍での実施ということで、遠隔授業を中心としたものとなった。対面での授業の必要性が議論される一方、それぞれの授業そのものが目指す到達目標については遠隔授業であっても到達できるものが多いことには異論はないと思う。統計

学やデータサイエンス分野におけるスキルセットの習得については、今回の事例においても示されたように、対面でなく遠隔授業形式でも十分であるといえる。一方、その活用力を十分に伸ばすためのグループワークの持ち方には、まだ課題も多く、今後事例を積み重ねていく必要がある。コロナ禍、コロナ後において、データサイエンス教育の重要性は増していきなかつ、遠隔授業を前提としたデータサイエンス教育の充実が望まれる。

参考文献

- 岩下志乃, 伊藤雅仁, 大野澄雄, 2015, 「JMOOC 講座を活用した反転授業の実施」『大学教育と情報』2015年度(1), 18-21.
- 花木良, 西仲則博, 伊藤直治, 近藤裕, 舟橋友香, 吉井貴寿, 2015, 「大学での数学学習における反転授業の導入に関する一考察」『次世代教員養成センター研究紀要』(1), 351-354.
- 渡辺美智子, 櫻井尚子, 山口和範, 井上達紀, 中川重和, 2000, 「総合情報処理教育としての統計分析能力の育成法とインターネット上の支援教材開発について」『情報教育方法研究』第3巻, 67-72.

URL

- AI 戦略 2019, <https://www.kantei.go.jp/jp/singi/tougou-innovation/pdf/aisenryaku2019.pdf> (2021年1月6日取得)
- GAISE, 2005, Guidelines for Assessment and Instruction in Statistics Education (GAISE) Reports, <https://www.amstat.org/asa/education/Guidelines-for-Assessment-and-Instruction-in-Statistics-Education-Reports.aspx> (2021年1月6日取得)
- JGSS2010, <https://ssjda.iss.u-tokyo.ac.jp/Direct/gaiyo.php?eid=0850> (2021年1月6日取得)
- モデルカリキュラム, 2020, http://www.mi.u-tokyo.ac.jp/consortium/model_literacy.html (2021年1月6日取得)

Summary

Statistical Education in Corona Disaster: Improvements based on the on-demand statistics courses at CSI

Seiichi Yamaguchi
Kazunori Yamaguchi
Minoru Kadota

With the global spread of COVID-19, university education in the year 2020 became centered on distance learning. Although face-to-face lectures are important, on-demand learning has great advantages in the acquisition of data analysis skills, in that learners can watch and check the lectures repeatedly at their own pace. The class size of a basic statistics class in a faculty of humanities and social sciences is often large, and the learners' basic skills in mathematics and computer skills vary widely. On the other hand, in learning to develop statistical thinking skills, it is important to provide opportunities for communication using evidence such as statistics. However, it becomes difficult to provide such opportunities in an on-demand format. In this paper, we introduce the case of on-demand distance learning of statistics at Rikkyo University in 2020.

Key words: Data science, Distance education, On-demand lecture