

講義コード	1200010401
講義名	データサイエンス
開講期	2024年度後期
⑤単位数	2単位
②授業形態	講義
⑥担当教員	田近 一郎
ナンバリングコード	22M2Z4201

③科目概要	社会現象・自然現象に関するビッグデータをインターネットを通して収集・抽出し、それらの現象に潜む複数の要因間の関係を分析して現象の理解や意思決定に活用する方法について学ぶ。その分析手段としての多変量データ解析の基礎である相関係数をはじめとして単回帰分析・重回帰分析、さらに因子分析、コレスポンデンス分析、ロジスティック回帰分析を取り上げ解説する。また、各種分析手法の実践として、インターネット上のさまざまなデータを収集・抽出し、解析・評価するまでの作業をクラウド開発環境Google Colab上でのPythonプログラミングによりおこなう。
①達成目標	各種の多変量データ解析手法を理解できる、実際のデータに適用する場面で適切な解析手法を選択して分析する技能を身につける。
履修に必要な予備知識や技能（関連科目等）	関連科目に「統計学」「知識情報学」がある。「基本情報技術者試験」で出題される散布図・回帰分析、データマイニングの分野の技能の習得を含む。
学位授与方針との関連	1. 情報メディアに関して、基礎的な情報処理技術、ソフトウェアについての知識、また、資格試験の学修内容等、基礎的な知識・技能を身につけている 2. 専修コースにおいて、情報理論、映像史、音響、色彩など、必要な知識・教養を身につけている。 5. 専門科目の学習の基盤となる教養と基礎教育の能力を身につける。
④授業計画、授業外学習の内容及び必要な時間	
第1回	多変量データと多変量データ解析の基本的な概念、適用事例、統計学全体における多変量データ解析の位置づけを解説する。また、本講義の範囲・目的、関連科目、講義の進め方や成績評価方法等も説明する。講義前、ビジネスにおけるデータサイエンスの位置づけをインターネット等で確認しておく（120分）。講義後、「量的データ」「多変量」で検索しそれらの意味について概略を把握する。またデータサイエンスの分析目的に関するWebClass課題を解く（120分）。
第2回	多変量データ解析が複数の変数間の因果関係と相関関係に注目する解析手法であることを解説し、その代表的な手法、そのプロセスを紹介する。クラウド開発環境Google Colabを起動し、プログラミング言語Pythonによる初歩的なプログラミングを通して一通りの使い方を学ぶ。講義前、インターネット等でさまざまな多変量データを収集しておく（120分）。講義後は、収集した個々のデータについて、データ中のどのパラメータ間で共変関係があるかを確認する。また因果関係と相関関係に関するWebClass課題を解く（120分）。
第3回	ピアソンの積率相関係数について、データの偏差、偏差積、共分散の意味を散布図を用いて図的に解説するとともに、相関係数を計算する手順について例を挙げて詳しく解説する。数理的な基礎知識としてΣ記号を用いる総和計算の各種具体例について解説する。講義前、「相関」等で検索し、相関に言及している分析例を下調べしておく（120分）。講義後、データの偏差、偏差積、共分散の意味についてまとめておく。また具体的なデータを対象に分散までの計算をWebClass課題としておこなう（120分）。
第4回	クラウド開発環境Google Colab上でPythonの変数・代入、条件分岐、繰り返し等の基本文法構造に関するプログラミング演習をおこなう。また、データを扱う際に多用するリスト、辞書等のデータ構造についても演習し、相関係数を求める簡単なデータ処理プログラムを作成する。講義前、プログラミング言語Pythonについてインターネット上で調査し、どのような目的で使われているかをまとめる。講義後、作成したデータ処理のためのPythonプログラムを調べ、相関係数を計算するための手順が実装されていることを確認する。また具体的なデータを対象に相関係数までの計算をWebClass課題としておこなう（120分）。
第5回	一つの原因から結果を予測する単回帰分析について解説する。単回帰分析の目的、手順を説明した後、例を挙げて2次元散布図で示された例から単回帰直線を実際に作る作業を行う。また、単回帰直線のグラフを作成する作業も行う。講義前、キーワード「単回帰分析」等で検索し、単回帰分析の具体例を下調べしておく（120分）。講義後は、その具体例の独立変数、従属変数等を把握し、単回帰分析の構造についてまとめる。（120分）。
第6回	クラウド開発環境Google Colab上でPythonの科学計算ライブラリNumpy、Scipyをインポートし配列・行列データの処理プログラミング演習をおこなう。次にデータ加工ライブラリPandasをインポートし汎用性の高いDataFrameを扱うプログラミングをおこなう。さらにグラフ描画ライブラリMatplotlibを利用してインターネットから取得したデータに対し散布図の作成と単回帰分析をおこなう。講義前、キーワード「DataFrame」等で検索し、データサイエンス用プログラムにおけるDataFrameの使用例を下調べしておく（120分）。講義後、単回帰分析の処理手順を復習する。また単回帰直線の導出に関するWebClass課題を解く（120分）。
第7回	複数の原因から結果を予測する重回帰分析を概説する。単回帰分析と重回帰分析の違いを説明した後、回帰平面の最小二乗法による導出方法について解説する。特に重回帰平面の係数の意味を詳しく解説する。講義前、キーワード「重回帰分析」等で検索し、重回帰分析の具体例を下調べしておく（120分）。講義後、その具体例の独立変数、従属変数等を把握し、偏回帰係数の意味をまとめる（120分）。
第8回	重回帰平面における重相関係数、決定係数について詳しく解説する。また、データに内在する多重共線性の問題と対処方法について解説する。クラウド開発環境Google Colab上でPythonの機械学習ライブラリscikit-learnをインポートし、インターネットから取得したデータに対して3次元散布図の作成と重回帰分析、重回帰平面の描画をおこなう。講義前、単回帰分析と重回帰分析について第7回プリントで復習する（120分）。講義後、重回帰分析の処理手順を復習する。またGoogle Colab上で実際のビジネスデータを対象に重回帰分析をおこない回帰平面を含む3D散布図を求めて提出する（120分）。
第9回	データをできるだけ少ない要因で説明する統計手法である因子分析の準備として、相関係数行列を説明する。因子分析の特殊な例として相関係数行列だけからの分析方法を例を挙げて解説する。また偏差ベクトルだけからの分析方法についても例を挙げて解説する。講義前、キーワード「因子分析」等で検索し、因子分析の具体例を下調べしておく（120分）。講義後、講義中に取り上げた具体例で得られている共通因子の導出の過程をまとめる（120分）。
第10回	因子分析の手順を、相関係数行列の固有値分解、共通因子の個数の決定から、因子負荷行列と因子得点行列の導出、座標軸の回転による共通因子の解釈にいたるまで概説する。主成分分析と因子分析の関係についてもふれる。講義前、キーワード「固有値分解」「因子負荷行列」「主因子法」「バリマックス回転」等で検索し、因子分析の手順を下調べしておく（120分）。講義後は、分析手順をトレースし、その概略を把握する。また相関係数行列だけから分析できるデータを対象とする初歩的な因子分析に関するWebClass課題を解く（120分）。
第11回	マーケティング関連データに対する因子分析の適用事例を紹介する。それからクラウド開発環境Google Colab上でPythonの機械学習ライブラリscikit-learnをインポートし、インターネットから取得したマーケティング関連データに対して因子分析をおこなう。講義前、因子分析の分析手順を第10回プリントで復習しておく（120分）。講義後、因子分析のプログラム上の処理手順を第10回のプリントを手掛かりに復習する（120分）。
第12回	アンケートの自由記述データの処理など質的変数間の内部構造を分析する対応分析（コレスポンデンス分析／数量化III類）をアンケートデータの例で解説する。講義前、キーワード「コレスポンデンス分析」「数量化」等で検索し、コレスポンデンス分析を適用した具体例を下調べしておく（120分）。講義後、その具体例のカテゴリスコア散布図、サンプルスコア散布図における第一軸、第二軸の解釈に至る手順を確認する。またクロス集計表だけから分析可能なデータを対象とする対応分析に関するWebClass課題を解く（120分）。
第13回	クラウド開発環境Google Colab上でPythonの機械学習ライブラリmcaをインポートし、インターネットから取得したアンケートデータに対してコレスポンデンス分析（数量化III類）をおこなう。講義前、コレスポンデンス分析の分析手順を第12回プリントで復習しておく（120分）。講義後、コレスポンデンス分析のプログラム上の処理手順を第12回プリントを手掛かりに復習する（120分）。

第14回	カテゴリ変数、量的変数が混合したアンケートデータの分析手法の一つであるロジスティック回帰分析について簡易的なアンケートへの適用例とビジネスでの本格的な適用例を用いて解説する。またロジスティック関数が指数関数を用いており確率と解釈できるように構成されていることを確認する。講義前、キーワード「ロジスティック回帰分析」で検索し、ロジスティック回帰分析の具体例を下調べしておく（120分）。講義後、「単回帰分析」「重回帰分析」「因子分析」「重回帰分析」「数量化IⅢ類」「ロジスティック回帰分析」について復習し、それぞれの分析手法の違いを扱うデータの種類の観点からまとめる（120分）。			
第15回	クラウド開発環境Google Colab上でPythonの機械学習ライブラリscikit-learnをインポートし、インターネットから取得したアンケートデータに対してロジスティック回帰分析をおこなう。多変量データの3次元散布図とそのデータをロジスティック回帰分析して得られたロジスティック曲面を重ねて表示しロジスティック曲面がデータの特徴を適切にとらえていることを確認する。講義期間中に出したミニプログラミング演習、ミニレポートの解説をおこなう。これまでに配布した授業プリントや補助資料を持参すること。欠席回の授業プリント等を把握した上ですべての授業プリントを整えること。講義前、提出したWebClass課題レポートを確認し理解不足の箇所について把握する（120分）。講義後、講義内容を復習し理解不足の内容に関するWebClass課題レポートを解く（120分）。			
講義進行方法、課題へのフィードバック方法	毎回配布する講義プリントへの記入とpdf補助資料、参考webサイトの紹介などで講義を行う。また、適宜講義内で紹介したデータ解析手順に関する演習を実施する。多変量解析上の重要な概念や用語は、その定義や説明を履修者が記入してゆく。データ解析手順を解説した後、適宜理解度の確認と習熟のためのPCによるミニ演習およびミニレポートを実施する。また、基本情報処理技術者試験などで実際に出題された問題もミニ演習の一部に組み込み、履修者に解いてもらう。提出されたサンプルコード等は動作に関して一定の品質を満たすまで修正作業を促す。フィードバックはレポート等は誤答の箇所を訂正し、適切な考え方を追記して返却する。			
アクティブラーニング	実習、フィールドワーク			
講義前・講義後の自主的学修活動への助言	講義前には学修のポイントやキーワードに従って、事前に下調べをしておき講義対象の把握に務めてもらいたい。また、講義後は、指示されたまとめや計算手順やデータ処理手順のトレースなどをおこない講義内容、処理手順の定着に努めてもらいたい。講義内容についての疑問点もそのままにせず教員に積極的に質問してもらいたい。			
⑦成績（達成度）評価方法・評価基準、割合	成績評価は、1. 定期試験（40%）、2. ミニ演習の達成状況等（35%）、3. ミニレポートの提出・達成状況等（15%） 4. その他（10%）に基づきおこなう。また、講義およびミニ演習やミニレポートの作業への取り組みの状況、講義前・講義後の自主的な学修活動に基づく積極的な質問等も評価する。			
教員の実務経験と授業科目との関連				
テキスト				
書名	著者	出版社	ISBN	備考
『なし』				
参考図書など				
書名	著者	出版社	ISBN	備考
『多変量解析のはなし-複雑さから本質を探る』	大村平	日科技連	978-4817180278	2006
『らくらく図解アンケート分析教室』	菅民郎	オーム社	978-4274066931	2007
『ビジネス活用事例で学ぶデータサイエンス入門』	酒巻隆治他	SBクリエイティブ	978-4797376333	2014
『東京大学のデータサイエンティスト育成講座 Pythonで手を動かして学ぶデータ分析』	塚本邦尊他	マイナビ出版	978-4839965259	2019
参考URL				
NO	表示名	URL	説明	
1.	統計学自習ノート	<a href="http://aoki2.si.gunma-u.ac.jp/lecture/index.html">http://aoki2.si.gunma-u.ac.jp/lecture/index.html</a>	統計学・データサイエンスの基礎から応用までを丁寧に幅広く取り上げている青木繁伸氏によるポータルサイト	