

統計検定 CBT「データサイエンス発展」 出題範囲表

大項目	小項目	ねらい	項目(学習しておくべき用語)
社会におけるデータ・AI 利活用	社会で起きている変化	社会で起きている変化を知り、数理・データサイエンス・AIを学ぶことの意義を理解する。	<ul style="list-style-type: none"> ・ビッグデータ、IoT、AI、ロボット ・データ量の増加、計算機の処理性能の向上、AIの非連続的進化 ・第4次産業革命、Society 5.0、データ駆動型社会 ・複数技術を組み合わせたAIサービス ・人間の知的活動とAIの関係性 ・データを起点としたものの見方、人間の知的活動を起点としたものの見方
	社会で活用されているデータ	社会でどのようなデータが集められ、どう活用されているかを知る。	<ul style="list-style-type: none"> ・調査データ、実験データ、人の行動ログデータ、機械の稼働ログデータ ・1次データ、2次データ、データのメタ化 ・構造化データ、非構造化データ(テキスト、画像/動画、音声/音楽) ・データ作成(ビッグデータとアノテーション)
	データ・AIの活用領域	さまざまな領域でデータ・AIが活用されていることを知る。	<ul style="list-style-type: none"> ・データ・AI活用領域の広がり(生産、消費、文化活動) ・研究開発、調達、製造、物流、販売、マーケティング、サービス ・仮説検証、知識発見、原因究明、計画策定、判断支援、活動代替、新規生成
	データ・AI利活用のための技術	データ・AIを活用するために使われている技術の概要を知る。	<ul style="list-style-type: none"> ・データ解析: 予測、グルーピング、パターン発見、最適化、シミュレーション・データ同化 ・データ可視化: 複合グラフ、2軸グラフ、多次元の可視化、関係性の可視化、地図上の可視化、挙動・軌跡の可視化、リアルタイム可視化 ・非構造化データ処理: 言語処理、画像/動画処理、音声/音楽処理 ・特化型AIと汎用AI、今のAIで出来ることと出来ないこと、AIとビッグデータ ・認識技術、ルールベース、自動化技術
	データ・AI利活用の現場	データ・AIを活用することによって、どのような価値が生まれているかを知る。	<ul style="list-style-type: none"> ・データサイエンスのサイクル(課題抽出と定式化、データの取得・管理・加工、探索的データ解析、データ解析と推論、結果の共有・伝達、課題解決に向けた提案) ・流通、製造、金融、サービス、インフラ、公共、ヘルスケア等におけるデータ・AI利活用
	データ・AI利活用の最新動向	データ・AI利活用における最新動向(ビジネスモデル、活用例)を知る。	<ul style="list-style-type: none"> ・AI等を活用した新しいビジネスモデル(シェアリングエコノミー、商品のレコメンデーション) ・AI最新技術の活用例(深層生成モデル、敵対的生成ネットワーク、強化学習、転移学習)
	データ・AI利活用における留意事項	データ・AIを扱う上での留意事項	データ・AIを利活用する上で知っておくべきこと
	データを守る上での留意事項	データを守る上で知っておくべきこと	<ul style="list-style-type: none"> ・情報セキュリティ: 機密性、完全性、可用性 ・匿名加工情報、暗号化、パスワード、悪意ある情報搾取、不正アクセス行為の禁止、個人認証、個人識別符号、要配慮個人情報、再識別、秘密の曝露や差別の誘引 ・情報漏洩等によるセキュリティ事故

データリテラシー	データを読む	データを適切に読み解く力を養う。	<ul style="list-style-type: none"> データの種類、データの分布と代表値、データのばらつき、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、外れ値 相関と因果(交絡、偏相関係数)、回帰(重回帰分析、ロジスティック回帰分析、モデルの評価) 分類とグループ化(階層的クラスタリング、非階層的クラスタリング) クロス集計表、分割表、相関係数行列、散布図行列 母集団と標本抽出(層別抽出、多段抽出、クラスター抽出、母数と統計量の区別、標本分布) 統計情報の正しい理解(誇張表現に惑わされない)
	データを説明する	データを適切に説明する力を養う。	<ul style="list-style-type: none"> データの表現(散布図、ヒートマップ、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、ランダム化比較試験) 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) 色の効果や特徴、点の色・大きさ・形状への配慮、線の太さと様々な破線
	データを扱う	データを扱うための力を養う。	<ul style="list-style-type: none"> 集計、並べ替え、ランキング、データ解析ツール(スプレッドシート)
数理基礎	線形代数	データ分析に必要なベクトルや行列の扱いやn次元ユークリッド空間の基本事項を理解する。	<ul style="list-style-type: none"> 平面ベクトル、空間ベクトル、n次元ベクトル、ベクトルの和、内積、直交性、ノルム 正方行列、単位行列、転置行列、対称行列、行列の積、逆行列、行列式 線形独立、部分空間
	微分積分	データ分析に必要な初等関数や微分積分の意味と操作を理解する。	<ul style="list-style-type: none"> 指数関数、対数関数、三角関数 積の微分、合成関数の微分、関数の最大最小、線形近似、原始関数、積分と微分の関係 偏微分、接平面、重積分、累次積分
	数列	数列の基本的な事項を理解する。	<ul style="list-style-type: none"> 数列の和、Σ記号、極限
デジタル情報とコンピュータの仕組み	デジタル情報	デジタル情報の表し方を理解する。	<ul style="list-style-type: none"> 数と表現(2進数の表現、論理値) 情報量の単位(ビット、バイト、接頭語(k,M,G,T,m,μ,n,pなど)を使った表現) 文字の表現(ASCIIコード、シングルバイト文字、ダブルバイト文字) デジタル化(連続値、離散値)、画像・動画(ラスタデータ、ベクタデータ、コーデック)
	コンピュータの仕組み	論理演算や計算上の誤差について理解する。	<ul style="list-style-type: none"> 集合、命題、真/偽、否定、論理和、論理積 有効数字、浮動小数点、仮数部、指数部、丸め誤差
アルゴリズム基礎	アルゴリズムの表現	アルゴリズムの表現方法を理解する。	<ul style="list-style-type: none"> フローチャート、アクティビティ図、端子、処理、判断、矢印
	アルゴリズムの構造	分岐、繰り返しなどのアルゴリズムの構造の基礎を理解する。	<ul style="list-style-type: none"> 代入、順次構造、選択構造、繰り返し構造
	基本的なアルゴリズムの例	いくつかの基本的なアルゴリズムを理解する。	<ul style="list-style-type: none"> 並べ替え(ソート)、探索(サーチ)、合計、併合
データ構造とプログラミング基礎(主にPython)	データ構造	配列などのデータ構造について理解する。	<ul style="list-style-type: none"> 配列とリスト(メモリ、ベクトル、行列、アドレス) 連想配列(キー、バリュー、連想、辞書、ハッシュ)
	プログラミング基礎	インタープリタ言語を用いて簡単なプログラミングができる。	<ul style="list-style-type: none"> インタープリタ言語(ソースコード、機械語、実行) 構文(変数、代入、計算、分岐、繰り返し)、演算(オブジェクト、四則演算) 関数(引数、返り値)、制御文(for, while, if文)、入出力(print文)
データハンドリング	代表的なデータ形式	代表的なデータ形式を理解する。	<ul style="list-style-type: none"> csv、XML、JSON
	その他のデータ形式	その他のデータ形式を理解する。	<ul style="list-style-type: none"> 離散グラフ、キー・バリュー形式である隣接リスト、NoSQL
	データベース	データベースの基礎概念を理解する。	<ul style="list-style-type: none"> データベース管理システム(DBMS)、リレーショナルデータベース、正規化、選択、射影、結合、SQL
	データクレンジング	データクレンジング作業を理解する。	<ul style="list-style-type: none"> 表記の揺れの吸収(文字列、数字、日付、時刻)、名寄せ
	データ加工	データの加工法を理解する。	<ul style="list-style-type: none"> 部分集合の抽出、行の並べ替え、新しい列の追加、プログラミング(Python、R)

データ取得とオープンデータ	日本や世界のオープンデータ	オープンデータの普及に向けた国内及び国際的な動きを理解する。	・二次利用可能なルール、機械判読への適性、オープンデータ憲章
	オープンデータの取得	オープンデータの取得法を理解する。	・e-Stat、DATA.GO.JP、データカタログサイト、Open Knowledge International、Web API
	統計法	統計法の意義について理解する。	・基幹統計調査、調査票情報の二次的利用
確率と確率分布	順列と組合せ	場合の数の数え方を理解する。	・階乗($n!$)、順列(mPn)、組合せ(mCn)
	確率分布の概念	確率変数の分布の基本を理解する。	・確率変数、確率関数、確率密度関数、母平均、母分散、同時分布、周辺分布、共分散と相関、独立
	主要な確率分布	主な確率分布と確率計算を理解する。	・ポアソン分布、指数分布、一様分布、正規分布、2変量正規分布
統計的推測	統計的モデル	統計的モデルの考え方を理解する。	・統計的モデル、母数、パラメータ
	標本分布	標本分布の基本的な考え方を理解する	・独立同一分布、標本平均、標本分散
	点推定	点推定について理解する。	・モーメント法、最尤法、バイアス、不偏推定量、平均二乗誤差、バイアス分散分解
	仮説検定の考え方	仮説検定の考え方を理解する。	・帰無仮説、対立仮説、2種の誤り、有意水準、検出力、p値、検定統計量
種々のデータ解析	時系列データ解析	時系列データの扱いを理解する。	・時系列データ(トレンド、周期、ノイズ)、季節調整、移動平均
	テキスト解析	テキスト処理の基礎を理解する。	・形態素解析、単語分割、ユーザ定義辞書、n-gram、文章間類似度、かな漢字変換の概要
	画像解析	画像解析の基礎を理解する。	・画像データの処理、画像認識、画像分類、物体検出
データ活用実践	教師あり学習	教師あり学習の実践例を理解する。	・教師あり学習による予測(例:売上予測、罹患予測、成約予測、離反予測) ・データの収集、加工、分析 ・データ分析結果の共有、課題解決に向けた提案
	教師なし学習	教師なし学習の実践例を理解する。	・教師なし学習によるグルーピング(例:顧客セグメンテーション、店舗クラスタリング) ・データの収集、加工、分析 ・データ分析結果の共有、課題解決に向けた提案

注: 統計検定3級、4級およびデータサイエンス基礎の範囲表の項目については、データサイエンス発展においても出題される。