

統計検定 CBT「データサイエンス基礎」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)	主な Excel 操作 (項目に渡るものは省略)
社会におけるデータサイエンス	社会におけるデータサイエンス	デジタル社会におけるデータサイエンスやビッグデータの役割、インターネットとデータサイエンスの歴史、個人データに関する情報倫理等を理解する。	インターネットとその歴史、Society5.0、問題解決、エビデンス、オープンデータ、官民データ活用推進基本法、データサイエンティスト、ビッグデータの特徴、機械学習手法の分類(教師あり、教師無し)、個人データ、情報倫理、AI(人工知能)手法の特徴、EBM(証拠に基づく医療)、EBPM(証拠に基づく政策立案)、IoT(モノのインターネット)	
データベース・データマネジメント	データベースマネジメント	分析目的に応じた構造化データの構築やデータ形式の変換、データ抽出等の簡単なデータの整理・整形ができる。	構造化データ(レコード×フィールド、ケース×変数)、欠測値、データの結合、データ形式(ロングフォーマット⇄ワイドフォーマット)、データ抽出(ランダムサンプリング、無作為標本抽出)、乱数	データのソート(並び替え) ピボットテーブル RAND 関数 データの分析 四則演算
	データマネジメント	データの種類や尺度を理解し、層別、水準(レベル)化、変数変換等のデータ処理ができる。	質的データ、量的データ、データの尺度、層別、水準(レベル)化、変数変換、Z 変換(標準化)、偏差値	IF 関数
データの可視化	データの可視化	データを目的に応じて可視化するための統計グラフの作成と解釈ができる。	円グラフ、棒グラフ、折れ線グラフ、帯グラフ、ツリーマップ、パレート図、ヒストグラム、箱ひげ図等	グラフの作成

質的データの分析	1変量の質的データの分析	質的データを用いて、問題の可視化や現状分析のためのパレート分析(ABC分析)ができる。	パレート表、パレート図、構成割合(確率)、累積度数(累積相対度数、累積確率)	SUM関数 ピボットテーブル
	2変量以上の質的データの分析	2つ以上の質的データを用いて、連関分析や要因探索のためのクロス集計表の分析ができる。	クロス集計表、行(列)比率、セル比率、期待度数とカイ 2 乗統計量、連関係数、特化係数、多重クロス表	CHISQ.TEST関数
量的データの分析	1変量の量的データの分析	量的データを用いて、問題の可視化や現状分析のためにデータの分布構造を分析できる。	階級、階級値、標準階級幅、度数分布表、ヒストグラム、基本統計量(平均、標準偏差、分散、四分位数、パーセント点)、箱ひげ図、変動係数、管理図、外れ値	データの分析 AVERAGE関数 VAR関数 STDEV関数
	2変量以上の質的・量的データの分析	2つ以上の質的データや量的データを用いて、要因探索のための分布の比較や相関分析、単(重)回帰分析による予測モデル構築ができる。	層別ヒストグラム、並列箱ひげ図、相関、相関係数、散布図、単回帰直線、重回帰モデル、寄与率、回帰係数、残差	CORREL関数
確率による意思決定	確率と確率分布	確率と確率分布による推測の考え方を理解し、シミュレーションを実行できる。	場合の数、確率、ベイズの定理、尤度、事後確率、期待値、2項分布、正規分布、確率的シミュレーション	BINOM.DIST関数 NORM.DIST関数 NORM.S.DIST関数
	推定	標本変動と誤差を理解し、母集団特性値の推定ができる。	信頼区間、信頼率、標本誤差、標準誤差 母平均、母比率	NORM.INV関数 Z.TEST関数
	検定	仮説検定の考え方を理解し、文脈に応じた検定を行い結果の適切な解釈ができる。	帰無仮説、対立仮説、有意水準(危険率)、有意確率(p値)、第一種の過誤、第二種の過誤、帰無仮説の棄却、2項検定、Z検定、t検定、 χ^2 検定、ABC分析テスト	T.TEST関数 CHI.DIST関数 CHI.INV関数 データの分析
時系列データの分析	時系列データの分析	時系列データの構造を理解し、特徴を分析できる。	指数、移動平均、伸び率、成長率	AVERAGE関数
テキストマイニング	テキストマイニング	テキストマイニングの意味を知り、単語や品詞の出現頻度を分析できる。		