

統計検定 CBT「データサイエンスエキスパート」 出題範囲表

大項目	中項目	小項目	キーワード(学習しておくべき用語)
統計基礎	確率と確率分布	確率分布、確率変数	チェビシェフの不等式、積率、尖度、歪度、積率母関数
		主要な確率分布	対数正規分布、ガンマ分布、ベータ分布、超幾何分布、負の二項分布、多変量正規分布
		確率変数の漸近的性質	大数の法則、中心極限定理、確率収束、分布収束
	推測統計	標本分布	カイ二乗分布、標本平均と標本分散の独立性、t 分布、F 分布
		点推定、区間推定	一致性、有効性、信頼区間と信頼係数
		汎用的な検定	尤度比検定、ノンパラメトリック検定、ウィルコクソン検定、並べ替え検定
		種々の検定	一元配置分散分析、二元配置分散分析、交互作用、適合度検定
		多重比較	ボンフェロニ補正
	ベイズ理論	事前分布・事後分布	事前分布、共役事前分布、事後分布
		ベイズ的仮説検定	ベイズファクター、ベイズ判別(各カテゴリーの事後確率)
	計算統計	ブートストラップ	復元抽出、経験分布、リサンプリング
		サンプリング	擬似乱数、逆変換法、棄却法、マルコフ連鎖モンテカルロ法
		モンテカルロ積分	モンテカルロ積分、期待値や確率密度の正規化定数、分散減少法
		欠測値の扱い	EM 法
数学基礎	線形代数	行列	三角行列、直交行列、行列とその逆行列の積の可換性、基本変形、ランク、簡約な行列、トレース
		データ記述と線形代数	all-ones ベクトル、偏差ベクトル、2つの偏差ベクトルの内積、射影行列、回帰分析における予測値ベクトルと残差ベクトル
		固有値と固有ベクトル	対称行列の固有値、固有ベクトル、対称行列の対角化、スペクトル分解、二次形式と(半)正定値行列、特異値分解
		n 次元ユークリッド空間	n 次元空間上の点の表現、線形部分空間と基底・次元、行列のランクとその列空間の次元、同次方程式、係数行列、解空間、解の一意性、正規直交基底、シュミットの直交化、射影と直交成分
		数値計算と線形代数	LU 分解、QR 分解、反復法
	微積分	1 変数関数の微分法	極大・極小と導関数、テイラー展開、方程式の数値的解法、反復法、二分法、ニュートン法
		1 変数関数の積分法	広義積分、ガンマ関数、ベータ関数、数列の収束と積分の収束
		多変数関数の微分法	勾配、極値と偏導関数の関係、ヘッセ行列、テイラー展開、ニュートン法、ヤコビ行列、連鎖律(多変数関数の合成関数の微分)
		多変数関数の積分法	重積分、重積分(長方形領域)、累次積分、一般の領域での重積分(縦線領域、横線領域の重積分)、変数変換とヤコビアン、広義重積分、ガウス積分、極座標変換、正規分布の極座標変換による求積
		数値積分	台形則、シンプソン法
	最適化	連続最適化	最急降下法、ニュートン法、ラグランジュ乗数法、条件付き極値問題、凸関数(定義、ヘッセ行列の(半)正定値性との関係)、最適性条件、線形計画法
		離散最適化	組合せ最適化、ネットワーク最適化、ナップサック問題、巡回セールスマン問題

計算基礎	コンピュータの構成	コンピュータの構成	コンピュータの構成・動作・性能
	データ収集	デジタルデータ収集	エッジデバイス、センサーデータ、プロトコル、インターネットの仕組み、Web クローラー、スクレイピング、クライアント技術(SDK、API など)、通信技術(HTTP、FTP、SSH など)、クラウドコンピューティングの利用、地理情報システム(GIS)、サイバーセキュリティ、ハッシュ値
	データ表現とデータ構造	データ表現	画像の符号化、画素(ピクセル)、色の 3 要素(RGB)、音声の符号化、周波数、標準化、量子化、データの圧縮と効率化
		データ構造	グラフ、ネットワーク、木構造、二分木、ヒープ
	データベース	データベース	テーブル定義、E-R 図、主キーと外部キー、データ操作言語(DML)、データ定義言語(DDL)、データウェアハウス(DWH)、正規化、射影、結合、SQL
	アルゴリズムとプログラミング	アルゴリズム	バブルソート、ヒープソート、幅優先探索、深さ優先探索、メタ戦略、貪欲法、局所探索、分割統治法、動的計画法、再帰的アルゴリズム、フローネットワーク、ビッグ O 記法、入力データ量、計算時間、ステップ数、最大回数、最急降下法、座標降下法
		プログラミング	計算の構造化、モジュール化、リファクタリング、オブジェクト指向、オブジェクト
モデリング・AI と評価	モデリング・AI による課題解決	AI の歴史と応用分野	AI の歴史、推論、探索、トイプロBLEM、エキスパートシステム、汎用 AI/特化型 AI(強い AI/弱い AI)、フレーム問題、シンボルグラウンディング問題、人間の知的活動と AI 技術(学習、認識、予測・判断、知識・言語、身体・運動)、AI 技術の活用領域の広がり(流通、製造、金融、インフラ、公共、ヘルスケアなど)
		モデル作成とデータ分析の進め方	現象のモデル化、仮説検証サイクル、分析目的の設定、様々なデータ分析手法、様々なデータ可視化手法、パターン発見、アソシエーション分析、リフト値、モデルの作成と検証、モデルの解釈と有効性、分析目的に応じた適切な調査(標本調査、標本誤差)、サンプルサイズの設計、ランダム化比較試験、実験計画法
	教師あり学習	線形回帰分析	正規方程式、回帰係数の有意性検定(t 検定、F 検定)、自由度調整決定係数、Stepwise 法、回帰分析の諸仮定の妥当性、多重共線性等の推測の信頼性
		質的回帰分析	ロジスティック回帰、オッズ比、対数オッズ
		判別分析	線形判別分析、二次判別分析、SVM、最大マージン判別、ソフトマージンと正則化、カーネル、非線形な分離曲面
		正則化法とモデル選択	バイアスとバリエーションのトレードオフの概念、リッジ回帰、lasso、スパースモデリング、ハイパーパラメータ、カーネル法
		決定木	決定木(回帰、判別)、アンサンブル学習(バギング、ランダムフォレスト)
		ベイズ統計・モデリング	単純ベイズ、ベイズ判別(各カテゴリーの事後確率)、階層ベイズ
	教師なし学習	クラスター分析	k-平均法の手続き、データの類似度、階層クラスタリングの手続き、最短距離法、最長距離法、Ward 法、樹形図(デンドログラム)
		主成分分析	可視化、特異値分解、寄与率、次元削減
		カーネル密度推定	ヒストグラム、カーネル関数
	その他の学習	強化学習	強化学習、Q 学習
	時系列解析	時系列データの特徴	周期性、強定常性、弱定常性、自己相関、相互相関、スペクトル
		時系列モデル	自己回帰(AR)モデル、自己回帰移動平均(ARMA)モデル、ARIMA モデル、SARIMA モデル、状態空間モデル、カルマンフィルタ
	生存時間解析	生存時間データ	打ち切り、生存関数、ハザード関数
		生存関数の推定	カプラン・マイヤー法、指数分布、ワイブル分布
	質的データ解析	質的データの解析	多重分割表、数量化理論、対応分析

	テキストデータ解析	テキストデータの数値化	形態素解析、tf-idf、ステミング、ストップワード、単語埋め込み、ベクトル空間モデル
		テキスト分析	共起ネットワーク、対応分析、トピックモデル、言語モデル、カナ漢字変換、機械翻訳
	モデルの評価	モデル評価指標	Mallows の Cp 基準、AIC、情報量規準、BIC、交差検証法、周辺尤度、MSE、正解率(accuracy)、適合率(precision)、再現率(recall)、混同行列、F 値(F1 値)、ROC 曲線とその AUC
		訓練データとテストデータ	訓練データとテストデータ、バイアスとバリエーションの概念、汎化誤差、過学習、適合不足、交差検証法(leave-one-out、k-Fold)、ハイパーパラメータ
	因果推論	因果モデル	実験研究と観察研究、潜在的結果変数、個体処置効果と平均処置効果、処置割当ての無視可能性、傾向スコア、マッチング、層別、重み付け法、ノンコンプライアンス
		グラフィカルモデリング	偏相関行列、距離行列、類似度行列、隣接行列、無向グラフ、有向グラフ、マルコフ確率場、ベイジアンネットワーク、構造方程式モデリング
	深層学習・ニューラルネットワーク	ニューラルネットワークの仕組み	入力層、出力層、シナプス結合、隠れ層、誤差逆伝搬法、勾配消失、活性化関数(ReLU、シグモイド関数、動径基底関数、ウェーブレット)、ドロップアウト、(確率的)勾配降下法とそのアルゴリズム、バッチ正規化、計算グラフ
		ニューラルネットワークモデル	CNN、フィルタ、プーリング、RNN、LSTM、画像解析、音声認識、転移学習
		生成 AI の基礎と展望	実世界で進む生成 AI の応用と革新、基盤モデル・大規模言語モデル・拡散モデル、生成 AI の注意事項、マルチモーダル(言語、画像、音声など)、プロンプトエンジニアリング、ファインチューニング、Transformer・注意機構、自己教師あり学習、敵対的生成ネットワーク(GAN)、Vision Transformer、CLIP、スケーリング則
	AI とロボット	AI とロボット	家庭用ロボット、産業用ロボット、サービスロボット、自動化機械、センサー、アクチュエータ、ジェスチャー認識、行動推定
	AI の構築・運用	AI の構築・運用	AI の学習と推論・評価・再学習、AI の社会実装・ビジネス/業務への組み込み、複数の AI 技術を活用したシステム(スマートスピーカー、AI アシスタントなど)、AI システムの開発・テスト・運用、AI システムの品質・信頼性、AI の開発基盤(大規模並列 GPU マシンなど)、AI の計算デバイス(GPU、FPGA など)

注: 統計検定2級、3級、4級、データサイエンス基礎、データサイエンス発展の出題範囲表の項目については、データサイエンスエキスパートにおいても出題される。この出題範囲表では、より高度な項目とそれらのキーワードのみを示している。