



データサイエンス系 資格検定まとめ

生成AI時代のDS検定のあり方は？

ChatGPTに聞いてまとめました

庭師

本資料で「検定」という言葉は
「資格検定」の意味で使っています。

図はChatGPT作成です。

はじめに：本発表の立ち位置

- ・ 庭師：自己紹介：

- ・ 大手製造業でデータサイエンスによる工程品質改善業務と新規事業企画業務をしています。社内Kaggleサークルもリードしています。

- ・ 本資料作成の動機：

- ・ Kaggleサークルで統計検定を受検し、体験発表する若手メンバーがいた。その前座としてデータサイエンス系検定の種類区分けとまとめをサークルメンバーに紹介するための資料を作成しました。
- ・ なのでKaggle Titanicスコアで理解しやすいようにしてみました。

- ・ 本資料の作成方法と正確性・客観性：

- ・ ほぼ全てChatGPTに問い合わせた資料にまとめました。正確性と客観性はChatGPTに依存します。
- ・  マークがついている問いをChatGPTへ入力しました。試してみてください。

- ・ 個人的見解としての発表です：

- ・ 特に何かを訴えたいわけではありません。「DS系検定をまとめていたら気になって、こんな質問をしてみました、紹介します」という立ち位置です。

一応、ファクト・チェックしましたが、ミスがあればご指摘を！



日本で主要なデータサイエンス系検定を表にまとめて

検定・資格名	主催	レベル感	特徴	主な対象者
統計検定®	日本統計学会・統計質保証推進協会	4級～1級、データサイエンス基礎	日本で最も有名な統計資格。統計理論から応用まで体系的。	学生、研究者、社会人実務者
データサイエンティスト検定™ リテラシーレベル	一般社団法人データサイエンティスト協会	初級～中級を対象としたリテラシーレベル	統計＋データ処理＋機械学習＋活用事例。2021年開始の新しい検定。	社会人・学生全般、文系も可
G検定(ジェネラリスト検定)	日本ディープラーニング協会(JDLA)	初級	AI/ディープラーニングの知識面。数式不要。業界動向や倫理も出題。	AIに関心のある社会人、文系含む
E資格(エンジニア資格)	日本ディープラーニング協会(JDLA)	上級	Python＋深層学習の実装スキルを問う。認定プログラム受講必須。	AIエンジニア、研究者
応用情報技術者試験	IPA(情報処理推進機構)	中級	IT全般知識。AI・統計も一部出題。 国家資格 。	ITエンジニア、社会人実務者

入門(文系・学生向け)

統計検定 3級・4級

データサイエンティスト検定

G検定

実務・社会人レベル

統計検定 2級

データサイエンティスト検定

応用情報技術者

専門家レベル

統計検定 準1級・1級

E資格



Kaggle titanicコンペでスコア0.80が出せる人には どの検定がお勧めですか？

- **Kaggle Titanic で LB 0.80 以上を出せるなら、すでに以下の力を持っていると考えられます：**
 - **pandas / scikit-learn / 可視化ライブラリを使った前処理・EDA能力**
 - **欠損値処理・特徴量エンジニアリングの発想力**
 - **モデル選択(ランダムフォレスト・GBDT・アンサンブルなど)**
 - **クロスバリデーションを使った評価スキル**
- **つまり「データ分析の実践力はすでに初級～中級を超えている」状態です。**

Kaggleレベル別おすすめ検定ルート

Kaggle Titanicレベル	特徴	おすすめ検定	狙い・効果
入門(Notebookを写経できる程度) LB ~0.70	・前処理・EDAの基礎習得中・モデルは写経で実行できる	- 統計検定 3級/4級- データサイエンティスト検定™ (基礎理解目的)	データの基本を体系化し、数学・統計リテラシーを固める
中級(自分で特徴量を作れる) LB ~0.75	・欠損補完や特徴量エンジニアリングを工夫できる・クロスバリデーションを理解	- 統計検定 2級- データサイエンティスト検定™	実務でも通じる「統計＋機械学習基礎力」を証明
上級(0.78～0.80に到達)	・アンサンブルやハイパラ調整ができる・モデル解釈や評価指標を理解	- 統計検定 準1級- G検定- 応用情報技術者試験 (AP)	データ分析＋AIリテラシー＋IT全般知識を横断的に習得
エキスパート(0.80超安定)	・特殊特徴量や論文知識を活用・他人のNotebookを超える工夫が可能	- 統計検定 1級(数理・応用)- E資格(実装力強化)	数理統計の裏付け＋深層学習実装力を武器に専門家レベルへ

- ・ 統計検定: 理論の深さを段階的に強化
- ・ データサイエンティスト検定: 幅広さの証明(実務・就活に)
- ・ G検定: AIのリテラシーをサクッと取得
- ・ E資格: 実装力を証明(AIエンジニア向け)
- ・ 応用情報: IT全般＋経営＋セキュリティの横断知識



今後、生成AIによるバيفコーディングが主流になるとしたら、データサイエンス系検定に意味はあるか？

バيفコーディング（AIが自動でコード生成・補助する時代）が本格化すると、従来の「人間が手でコードを書けるか」を測る試験の意義は揺らぎます。単なるスキル確認にとどまらない方向に検定内容がシフトしていくと考えられます。

DS資格検定で重要度が上がるもの

- **生成AI普及後に残る「検定の意義」**
- **基盤知識の保証**
 - AIがコードを書けても、仕組みが分からなければ「正しい指示」や「結果の検証」ができません。統計検定やデータサイエンティスト検定は、AIの出力を批判的に評価する基礎知識の証明として価値を持ち続けます。
- **リテラシーの証明**
 - G検定のような「AIリテラシー」資格はむしろ重要度が上がります。なぜならAIが生成する結果を社会・倫理・法の観点で適切に扱う力が求められるからです。
- **専門性の裏付け**
 - E資格や統計検定1級のように「数理的裏付け」や「実装理論」を問う検定は、AI時代でも専門家ポジションの区別に有効です。

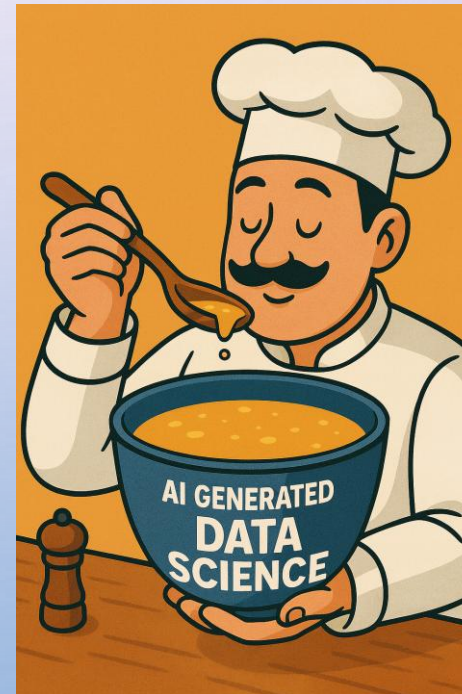
DS資格検定で重要度が下がるもの

- ・ 単純な暗記系・ハンスオン実装力のみを問う試験

生成AIで簡単に代替可能な「文法知識」「APIの呼び方」などは資格の意義が弱まります

→ 今後は「自動生成をどう使いこなすか」「生成物をどう検証・改善するか」に重点が移ります。

自動生成をどう料理
するか



自動生成をどう
検証評価するか

今後の方向性と結論

- 今後の検定の方向性

- AI活用前提の出題「AIが出力した結果の誤りを指摘せよ」「生成AIが提案した特徴量の妥当性を評価せよ」など、AIと人間の協働を前提とした試験に変わる可能性。

- 横断力の強化

- 応用情報技術者試験(IPA)のように、統計・AIに加えてネットワーク・セキュリティ・経営戦略まで含めた「広い視点」を持つことが価値になります。

- 結論

- 生成AIの時代でも、
 - 基礎理論を保証する検定(統計検定、E資格)
 - リテラシーを保証する検定(G検定、データサイエンティスト検定)
- は意味を持ち続けます。ただし、「コードが書ける」だけを測る試験の価値は下がり、AIを正しく使える力を示す資格の価値が上がる 方向にシフトしていくでしょう。



将来の「AI前提」検定のイメージ と問題例シミュレーション

出題の方向性

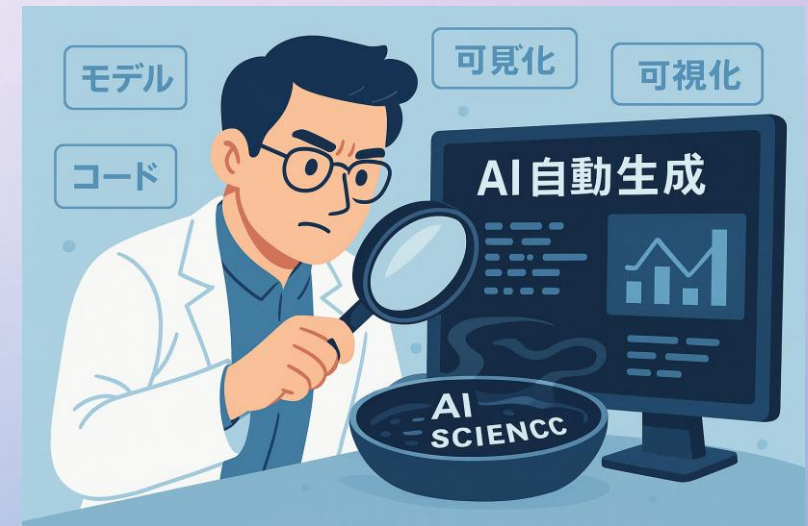
- AI出力の評価力:
- AIへの指示力(プロンプト設計力):
- エラーやバイアス検出力:
- 社会実装・倫理判断:

これで決定というわけでは無く、こういう方向性議論があいうる、的に見て頂ければと...

皆さんと良い議論ができれば幸いです。

問題例シミュレーション: AI出力の評価問題

- AIが生成したモデル・コード・可視化の妥当性を見抜く問題。
- 問題:
 - 下記は生成AIがTitanicデータセットを分析したレポートの一部です。誤りを2つ指摘し、修正方法を説明してください。
- AI出力の要約:
 - Age列の欠損値は「0」で補完した。
 - モデル評価には「学習データのAccuracy」を使用した。
- 解答例の方向:
 - 0補完はバイアスを導入 → 「中央値 / 回帰補完」が妥当。
 - Accuracyのみは不適切 → クロスバリデーションやF1を用いるべき。



問題例シミュレーション: フロント設計力問題

- 欲しい分析を適切にAIに依頼できるかを問う。
- 問題:
 - あなたは「顧客離反率を予測するモデル」をAIに作らせた。以下の指示文を改善しなさい。
- 元のフロント:
 - 「顧客データで予測モデルを作って」
- 解答例の方向:
 - 「CSV形式の顧客属性・契約履歴を用いて、二値分類モデル(離反:1 / 非離反:0)を構築してください。欠損値補完の方法も明示し、学習・検証データを8:2に分けてROC-AUCを評価指標としてください。」



問題例シミュレーション：エラーやバイアス検出問題

- AI出力の偏り・誤いを批判的に検証する力。
- 問題:
 - 生成AIが作ったクレジット審査モデルの重要特徴量ランキングに「性別」が含まれていました。
倫理・法務の観点から、どのようなリスクがあり、
どう対応すべきか述べなさい。
- 解答例の方向:
 - リスク: 差別的判断(ジェンダーバイアス)、法令違反の可能性。
 - 対応: 特徴量から性別を除外、代理変数の確認、フェアネス指標(Demographic Parityなど)を追加。



問題例シミュレーション：社会実装・倫理問題

- AIの適用に伴う倫理・法務・ビジネス上の影響を問う。

- **問題:**

- AIが自動生成した医療診断モデルを使う場合、責任の所在は誰にあるべきか？
選択肢から最も妥当なものを選べ。

1. AIベンダー
2. 病院の運営者
3. 医師本人
4. 法律で定める第三者機関

- **解答例の方向:**

- 「3. 医師本人」+「責任を補完する形でAIベンダー・病院も説明責任を負うべき」など複合的に論じる。



まとめ

- **従来: コード・統計式・暗記中心**
- **未来: AIをどう使うか／どう批判的に検証するか**
- **今後、問われる力:**
 - **AIリテラシー (G検定的要素)**
 - **データ検証力 (統計検定的要素)**
 - **AI実装＋評価力 (E資格的要素)**
 - **社会実装の理解 (応用情報的要素)**