国立大学法人
名古屋工業大学

第17回日本統計学会春季集会
03/04/23 @都立大 (20分)

統計検定CBTの試験問題評価とその課題

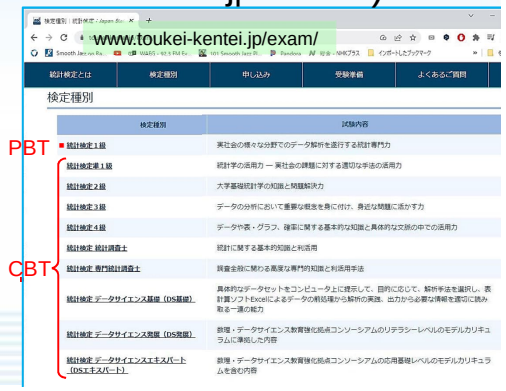
桜井 裕仁 (大学入試センター)
○ 林 篤裕 （名古屋工業大学)
本多 正幸 (千葉大学)

e-mail: hayashi.atsuhiro@nitech.ac.jp

資料掲載URL: stat.web.nitech.ac.jp/haifu/#JSS2303

1. 背景(1)

- 統計検定
  - 認定: 一般社団法人 日本統計学会
  - 実施: 一般財団法人 統計質保証推進協会
- 検定種別(https://www.toukei-kentei.jp/exam/)
  - 10種類:
    - 1級から DSエキスパートまで
  - 1級以外の9種類は全てCBT方式
  - 1級はPBTを継続



1. 背景(2)

- 各級の分科会
  - 統計検定の作成・運営: それぞれで会合
  - 合同で定期的に評価・分析に関する会合を開催
    - 各級の試験問題の評価等を意見交換・情報共有
- CBT問題評価分科会: 各級の分科会とは独立
  - CBTでの運用に関する検討や支援
- 本稿: 活動の紹介、今後の課題

2.1 資格試験に求められるもの(1)

- 資格試験:「資格付与の際、それにふさわしい力量を持っているかどうかを調べる試験」
  (広辞苑第七版)
- 内容(分野): 事前に定められた出題範囲の中から偏りなく出題
- 難易度: 広範に準備
  ＝受検者の勉強度合い(能力)に対応するため

2.1 資格試験に求められるもの(2)

- 例えば、2級の出題範囲
  - 統計検定のWebページ
    (https://www.toukei-kentei.jp/exam/grade2/)
  - 公式問題集
    (日本統計学会, 2023)
  - サブグループ
    - 大項目で10区分
    - 小項目で21区分

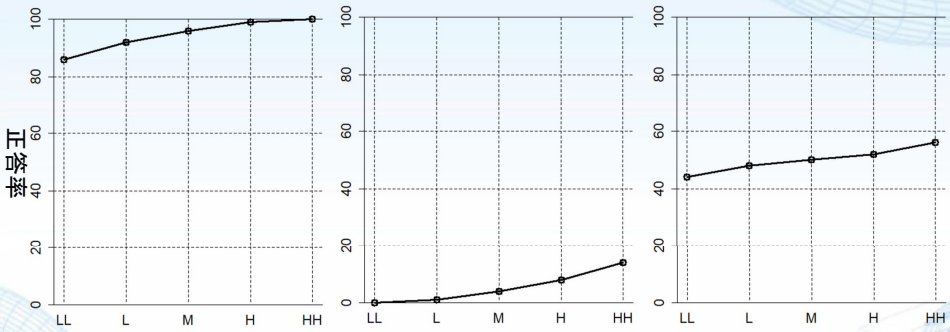
統計検定2級 出題範囲表

大項目	小項目	ねらい	項目(学習しておくべき用語)
データソース	身近な統計	歴史的な統計学の活用や、社会における統計の必要性の理解、データの取得の重要性を理解する。	(調べる場合の)データソース、公的統計
データの分布	データの分布の記述	集められたデータから、基本的な情報を抽出する方法を理解する。	質的変数(カテゴリカル・データ)、量的変数(数値)、度数分布表・ヒストグラム、歪度(分布の高い、対称、ベル型、一様、単峰、
1変数データ	中心傾向の指標	分布の中心を説明する方法を理解する。	平均値、中央値、最頻値(モード)
	散らばりなどの指標	分布の散らばりの大きさなどを評価する方法を理解する。	分散(n-1で割る)、標準偏差、範囲(最大値-最小値)、四分位範囲、シニ尖度、2つのグラフの視覚的比較、歪度、尖度
	中心と散らばりの活用	標準偏差の意味を知り、その活用方法を理解する。	偏差、標準化(z得点)、変数係数、指数
2変数以上のデータ	散布図と相関	散布図や相関係数を活用して、変数間の関係を探る方法を理解する。	散布図、相関係数、共分散、層別した散布図、相関係数
	カテゴリカルデータ	質的変数の関連を探る方法を理解する。	度数表、2元クロス表
データの活用	単回帰と予測	回帰分析の基礎を理解する。	最小二乗法、変動の分解、決定係数、散布図、ブロット、標準誤差、変数変換
	時系列データの処理	時系列データのグラフ化や分析方法を理解する。	成長率、指数化、移何平均、系列相関・
推測のためのデータ収集法	観察研究と実験研究	要因効果を測定する場合の、実験研究と観察研究の違いを理解する。	観察研究、実験研究、調査の設計、母集団、ランダム、無作為抽出
	標本調査と無作為抽出	標本調査の基本的概念を理解する。	標本サイズ(標本の大きさ)、標本誤差、層別抽出法、クラスター抽出法、多段抽出法
	実験	効果評価のための適切な実験の方法について理解する。	実験のデザイン(実験計画)、フィッシャー
確率モデルの導入	確率	推測の基礎となる確率について理解する。	事象と確率、加法定理、条件付き確率、
	確率変数	確率変数の表現と特徴(期待値・分散など)について理解する。	離散型確率変数、連続型確率変数、確率関数、和と差(同時分布、和の期待値・分散)、
	確率分布	基礎的な確率分布の特徴を理解する。	ベルヌーイ試行、二項分布、ポアソン分布、2変量正規分布、超幾何分布、集合
推測	標本分布	推測統計の基礎となる標本分布の概念を理解する。	独立試行、標本平均の期待値・分散、大数の法則、二項分布の正規近似、連続修正、
		正規母集団に関する分布とその活用について理解する。	標準正規分布、標準正規分布表の利用、上側確率点(パーセント点)
	推定	点推定と区間推定の方法とその性質を理解する。	点推定、推定量と推定値、有限母集団
		1つの母集団の母数の区間推定の方法を理解する。	正規母集団の母平均・母分散の区間推
		2つの母集団の母数の区間推定の方法を理解する。	正規母集団の母平均の差・母分散の比
	仮説検定	統計的検定の意味を知り、具体的な利用方法を理解する。	仮説検定の理論、p値、棄無仮説(帰無仮説)と対立仮説、第1種の過誤、第2種の過誤、検出力
		1つの母集団の母数に関する仮説検定の方法について理解する。	母平均の検定、母分散の検定、母比率
		2つの母集団の母数に関する仮説検定の方法について理解する。	母平均の差の検定(分散既知、分散未知でない場合)、母分散の比の検定、母比率
		適合度検定と独立性の検定について理解する。	適合度検定、独立性の検定
線形モデル	回帰分析	重回帰分析を含む回帰モデルについて理解する。	回帰直線の傾きの推定と検定、重回帰モデル、線形性、ダミー変数を用いた回帰、自由度
	実験計画の概念の理解	実験研究による要因効果の測定方法を理解する。	実験、処理群と対照群、反復、ブロック化、乱数分割し、ずれ
活用	統計ソフトウェアの活用	統計ソフトウェアを利用できるようになり、統計分析を実施できるようになる。	計算出力を活用できるか、問題解決に

2.2 試験問題の評価方法

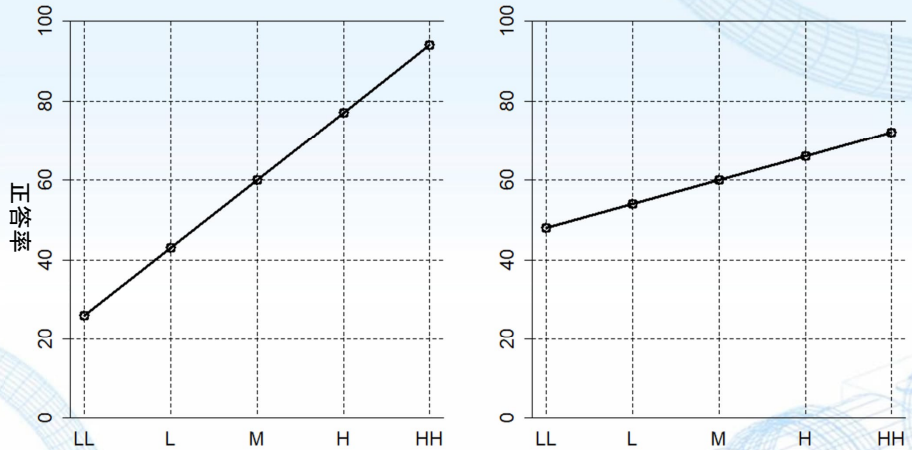
- 一般的に言われる“難易度”
  ＝受検者全員を対象とした 全設問の正答率(平均点)
- 試験の評価指標としての「難易度」・「識別力」
  - 受検者の勉強度合いに依って
    正答率が変動することに注目した評価指標
  - 設問解答率分析図(五分位図 or 三分位図)
    - 受験者を総合成績順に人数がほぼ均等な5群(3群)に分割
    - 「難易度」: 折れ線の上下位置
    - 「識別力」: 折れ線の傾き

典型的な例(図1)：難易度



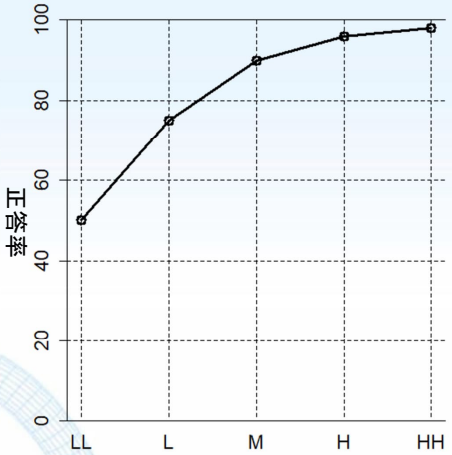
どの群にも易しい　どの群にも難しい　どの群にもそこそこ

典型的な例(図2)：識別力

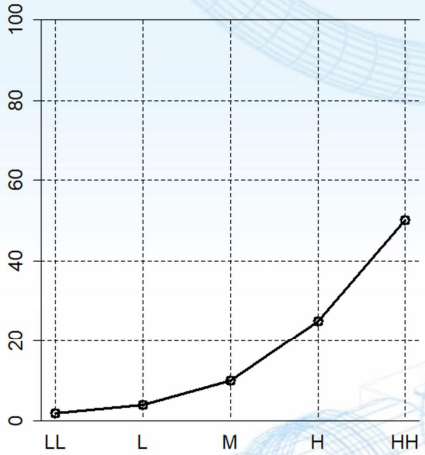


各群を分離し易い　各群を分離しにくい

典型的な例(図3)：部分的識別

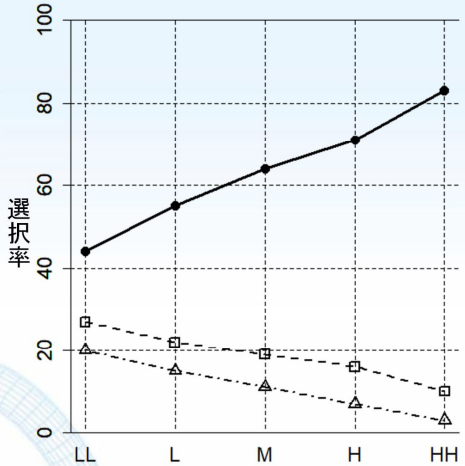


下位群を識別

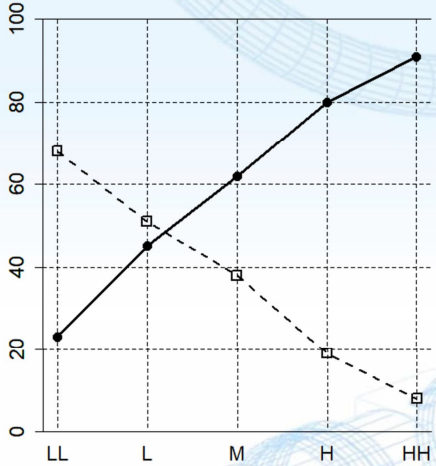


上位群を識別

誤答選択肢を追記(図4)



受験者の選択行動



2極化(2択)

分析図の性質

- 各群の勉強度合い(能力)に見合った正答率
  - 基本的には右上がりになるはず ➡ 単調増加
- 難易度：線分の高さ：図1
- 識別力：線分の傾き：図2、図3
  - 増加の程度、直線の勾配
  - 各群で正答率に差がある & 単調増加：識別に有効
  - 増加の程度が低い：識別には有効でな
- 誤答分析:惑わされ易い選択肢：図4

2.3 分析図の活用方法

- 設問解答率分析図：設問単位で分析
  - 正答選択肢: 単調増加が基本だが.....
    - 作題工程の見直しの動機付けに
  - 誤答選択肢: 受検者が魅力的に感じた選択肢
    - 将来の作題に活用
- 大問得点率分析図：大問単位で分析
  - 「設問」が幾つかずつにまとまって「大問」を構成
  - 大問というまとまりでの“正答率” ＝ “得点率”
    - 大問の場合には、構成する設問の配点が加味
  - 大問単位での難易度・識別力

2.4 CBT方式の使用に関して

- CBT方式
  - 予め設問を組み合わせた複数のセットを準備
    - 出題分野や難易度が偏らないように
- 出題範囲をいくつかのサブグループに分割し、サブグループごとに管理・組み合わせ
  - サブグループ内での組み合わせの吟味
  - サブグループ間での難易度や識別力の差異も評価
- 受検者群に依って設問や大問が異なる
  - 受検者群ごとに設問や大問の評価を行う必要
  - 大問間の相関や得点分布等に関する指標も算出

2.5 統計検定に向けての対応

- 受検者の解答状況
  - 適宜小セットに分割・結合しながら分析
    - 受験者群、設問群、大問群
  - 一つの級だけでも非常に緻密な作業が必要

    ===> 多大な労力

- CBT問題評価分科会
  - 分析結果をシステマチックに提示できるような方策を
  - CBTの運営委託会社である (株)オデッセイ コミュニケーションズと協力しながら

3. 今後の課題

- 喫緊の課題：評価体制の確立
  - 各種の分析結果をタイムリーに提供
  - 各級の分科会で行っている分析作業の支援
- 検討事項
  - 出題分野のサブグループ内でのばらつきの許容性
    - どのように慎重・丁寧に作題作業を行ったとしても生じる
  - 設問・大問のライフサイクル：見極め方の確立
    - ある程度曝露された設問・大問は入れ替えるフェーズ
      - 設問・大問は複数回の試験で提示される

Sustainable

- 統計検定を継続的・安定的に実施していくために
  - 評価資料の整備、効率的な検討環境の確立