

## 第1章 測定と評価：妥当性 (validity) と信頼性 (reliability)

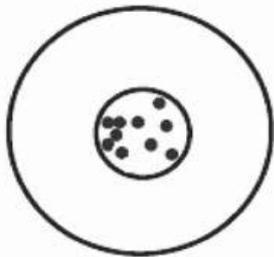
### 1-1 統計分析を行う前に

#### 1-1-1 測定と評価のための統計分析

教育・心理学分野の仮説検証型 (hypothesis testing) の研究では、適切な尺度 (scale、質問事項やテストなど) を使って評価 (evaluation) をする。

#### 1-1-2 測定の二大要素：妥当性と信頼性

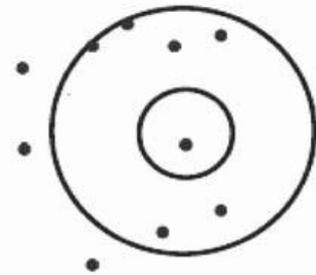
高い妥当性、低い信頼性



低い妥当性、高い信頼性



低い妥当性、低い信頼性



### 1-2 妥当性の捉え方と検証法

#### 1-2-1 伝統的な妥当性の分類

1966年、米国教育学会 (AERA)、米国心理学会 (APA)、全米教育測定会議 (NCME) の合同委員会による妥当性の概念の整理

表 1.1 伝統的観点から見た妥当性の分類

妥当性の種類	定義	検証法
内容的妥当性 (content validity)	項目が、どの程度全体を偏りなく代表しているか。	その分野の専門家に判断してもらう。
基準関連妥当性 (criterion-related validity)		
1. 併存的妥当性 (concurrent validity)	外部の基準とどの程度関連があるか。	同時期に得た基準との相関を検証する。
2. 予測的妥当性 (predictive validity)	その尺度の実施以降の変化などをどの程度適切に予測しているか。	尺度の得点とその後の基準との相関を検証する (第7章)。
構成概念妥当性 (construct validity)	測定しようとする構成概念をどの程度適切に反映しているか。尺度の得点とモデルや理論との整合性がどれだけあるか。	信頼できる基準との相関や因子分析 (第9章), MTMM (収束的・弁別的妥当性: 2-3 参照) などによって検証する。

predictive validity ☹ content validity ≈ construct validity ☹

1-2-2 新しい妥当性の定義: Messick [1995, 1996] の分類

表 1.2 構成概念妥当性の6つの側面

妥当性の側面	妥当性の証拠	具体的な証拠の収集法
内容的	項目の内容が設定したドメインの内容に対応しているか、十分にドメインを代表しているかを示す証拠。	・カリキュラム分析, タスク分析, 談話分析など ・専門家や尺度のユーザーによる判断
本質的	項目やタスクなどの反応 (回答) プロセスが心理学的に理論的根拠があるという証拠。	・タスク中の発話プロトコルやタスク後の再生 ・目の動きや反応時間などの観察 ・質問紙・インタビュー
構造的	尺度内の項目間の関係が理論的構造に一致していることを示す証拠。尺度の次元性や応答パターンが心理測定モデルに合うという証拠。	・得点化手続きの適切性の診断 ・相関分析 (第7章) ・因子分析や構造方程式モデリング (第9, 10章) ・項目応答理論
一般化可能性	当該データの理論的特性 (平均や標準偏差, 項目間の相関構造など) が一貫しており, かつ他の集団, 実施場面, 実施時期, 同様の項目セットに対しても不変であるという証拠。	・一般化可能性理論 ・再検査信頼性 (再テスト信頼性) (第7章) ・代替検査信頼性 (同等フォーム) (第7章) ・アルファ係数 (本章1-3-3)
外的	当該尺度と他の尺度の得点間に理論上想定される相関パターンが示されるという証拠。	・相関分析や MTMM (本章1-2-3) ・構造方程式モデリング (第10章)
結果的	その尺度を使用することの適切さ。短期的・長期的な悪影響が理論的・経験的に生じない, 予見されないという証拠。	・観察・質問紙・インタビューなどによる波及効果 (本章1-2-4) の分析 ・事前・事後テストの分析

1-2-3 多特性・多方法行列分析法 (multi-trait multi-method analysis: MTMM)

同じ受験者に対して2つ以上の特性を2つ以上の方法で測定し、相関行列分析を使って収束的妥当性 (convergent validity → monotrait-heteromethod correlation) と弁別的妥当性 (discriminant/divergent validity → heterotrait-monomethod correlation / heterotrait-heteromethod) を満たしているかで構成概念妥当性があるかどうか確かめる。

#### 1-2-4 波及効果

Messick の 6 つの側面のうちの一つの結果的側面に該当する。  
 テストが教育機関や社会に与える影響 (impact)。

### 1-3 テストの信頼性

#### 1-3-1 信頼性の定義

得点の安定性あるいは一貫性

#### 1-3-2 信頼性係数の推定方法

表 1.4 代表的な信頼性の推定方法

信頼性の推定方法	具体的な方法
再テスト法 (test-retest reliability)	同じテストを同一の受験者に 2 回実施し、その点数の一致度を相関係数で示す。2 回のテスト実施の間に被験者の能力が変わらないこと、テストの内容を被験者が忘れていないことが条件。2~3 週間の間隔を空けて行う。
同等フォーム法 (equivalent-forms reliability)	ほぼ同一の難易度、標準偏差をもつ 2 つのテスト (平行テスト) を同一の被験者に実施し、その点数の一致度を相関係数で示す。
内的一貫性 (internal consistency) ・折半法 (split-half method) ・アルファ係数 ( $\alpha$ coefficient; Cronbach's $\alpha$ ) ・キューダー・リチャードソン 20 (KR-20) ・キューダー・リチャードソン 21 (KR-21)	<ul style="list-style-type: none"> <li>たとえば、項目を奇数番号と偶数番号に分けて、この両者の点数の一致度 (相関係数) を計算し、スピアマン・ブラウン公式で修正して算出する (第 7 章)。</li> <li>項目の標準偏差とテストの総得点の標準偏差により算出する (第 7 章)。</li> <li>項目数、項目分散の合計、テスト全体の分散により算出する。2 値データのみ適用。</li> <li>テスト項目数、平均値、標準偏差により算出する。2 値データのみ適用。</li> </ul>
評価者信頼性 (rater reliability) ・評価者間信頼性 (inter-rater reliability) ・評価者内信頼性 (intra-rater reliability)	<ul style="list-style-type: none"> <li>評価者が 2 名の場合は、相関係数やカッパ係数 (第 7 章) を用いる。評価者が 3 名以上の場合、アルファ係数を使用する。また一般化可能性理論や項目応答理論を用いて算出することも可能。</li> <li>1 人の評価者が、同一のパフォーマンスをある一定の期間において複数回採点した結果が一致する度合い。算出方法は、評価者間信頼性に準ずる。</li> </ul>