

■測定信頼性

心理的構成概念の測定

その特性が高い人だったらこう答えるだろう、低い人だったらこう答えるだろうという区別がなるべくつくような項目を複数提示し、それらに対する反応によって、構成概念の程度を評価するという間接測定

→ 誤差を含む

誤差はないに越したことはない。

測定の誤差を評価する概念が必要。

信頼性：尺度（テスト）が実際に測定している特性をどの程度精度良く測定しているかを考える概念。テストの得点または各項目の得点が、各受験者において一貫している程度で評価される。

一貫性：もう一度同じテストをやったら同じ得点が得られるという一貫性（再現性）や、同じような項目の得点は同じような値になるという一貫性（内的整合性）などがある。

→ 信頼性の具体的な指標は1通りではなく、幾通りかの捉え方がある。

信頼性係数にはいくつか種類がある

← 信頼性の定義は1つ

誤差E：ヤマの当たり外れ、ど忘れ、勘違い、その日の体調、気分、読み違い、たまたま得意分野だった、そこだけは覚えていた（逆に抜けていた）など

真の得点T：誤差の影響がない、本当のその受験者の得点

観測得点X：実際の観測得点（データ）

テスト理論の基本モデル

$$X = T + E \quad \text{観測得点} = \text{真の得点} + \text{誤差}$$

実際のテスト得点は真の得点と誤差が足しあわされたものであると仮定する。

仮定

誤差の平均はゼロ：+の誤差の人、-の誤差の人いろいろいるだろうが、平均すればゼロになるという仮定。

真の得点と誤差は無相関：真の得点が高いほどヤマが当たりやすい（誤差が+に大きい）などのような関連はなく、真の能力がどんな程度でも、誤差は±に無相関に散らばる、という仮定。

これらの仮定から出てくる帰結

$$\text{観測得点の分散} = \text{真の得点の分散} + \text{誤差の分散}$$

実際の観測得点の散らばりは、真の得点の散らばりに、誤差の散らばりが加わったもの。

誤差がない → みな誤差ゼロ → 誤差の分散ゼロ

→ 誤差が少なければ、誤差の分散は小さくなる。

■信頼性係数の定義

信頼性係数 = 真の得点の分散 / 観測得点の分散

観測得点の散らばりうち、真の得点の散らばりは何割か、反対にいうと、実際のテスト得点に、誤差の影響はどれだけ入っていないかを表す指標

実際に観測できるのは観測得点の分散だけで、真の得点の分散や誤差の分散は観測できない。→ 信頼性係数の推定

■信頼性係数の推定

再検査信頼性係数：もう一度同じテストをやったら同じ得点を得られるという一貫性（再現性）に注目。同じテストを2回実施する。

再検査信頼性係数 = 2回の観測変数の相関係数

再検査の間隔は、真の得点Tが同じと考えられる時点で、経験的に2週間～2ヶ月程度が用いられている。

内的整合性信頼性係数：同じような項目の得点は同じような値になるという一貫性（内的整合性）に注目。同じ潜在変数と関連すると思われる項目を複数個提示する。

α 係数 = 項目数 / (項目数 - 1) × (1 - 各項目分散の和 / 合計得点の分散)

平成19年度調査の千葉県データにおける、各テストの信頼性係数（ α 係数）
小学6年生 約5万3千名 中学3年生 約4万7千名

教科	項目数	α 係数	教科	項目数	α 係数
小学校国語A	18	0.75	小学校算数A	19	0.84
小学校国語B	10	0.75	小学校算数B	14	0.76
中学校国語A	37	0.87	中学校数学A	36	0.94
中学校国語B	10	0.77	中学校数学B	17	0.87

国語のテストより算数・数学のテストの方が α 係数は高くなり安いという一般的傾向がある。

一般に α 係数は、項目数が多くなると値が1に近くなる傾向があるという性質を持つ。中学校国語A、数学Aの α が高いのは、その影響もあるだろう。

項目数を多くすると疲労効果など別の影響が出てくる → 妥当性の問題