

⇒ 信頼性・妥当性の双方が高いこと
両者は同じではない

信頼性(reliability)が高い

- ・ 再現性→いつ、どこで、どのように測定しても同じ結果が得られるか
- ・ 等質性→テスト項目が同じ側面を測っているか
- ・ 信頼性係数で表現

妥当性(validity)が高い

- ・ 測定しようとするものを的確に測定しているか
- ・ 妥当性係数で表現できる場合あり

○再テスト法

- 同一テストの繰り返し実施。再現性・安定性
- 信頼性係数(相関係数)

× 平行テスト法

× 折半法

- 信頼性係数(スピアマン・ブラウンの公式)

◎内部一貫法(内的整合法)

- テスト項目の間関係の強さ。等質性・識別性
- 信頼性係数(α 係数、KR-20の式 など)

◎一般化可能性理論による方法

- 分散成分の推定による

再テスト法

- 同一対象に同一テストを複数回行う
- 得点の間で相関係数(r)を求める

内部一貫法(内的整合法)

- テスト項目同士の関係が強いかどうかによる

$$\alpha = \frac{\text{項目数}}{\text{項目数} - 1} \times \left(1 - \frac{\text{項目ごとの分散の合計}}{\text{合計点の分散}} \right)$$

内的整合法(内部一貫法)の α 係数の 計算法と考え方

$$\alpha \text{ 係数} = \frac{\text{問題数}}{\text{問題数} - 1} \times \left(1 - \frac{\text{問題項目の分散の合計}}{\text{個人得点の分散}} \right)$$

番号	問1	問2	問3	問4	問5	個人得点
1	2	2	2	2	1	9
2	2	2	1	2	2	9
3	2	2	1	2	2	9
4	2	2	2	2	1	9
5	2	1	0	2	2	7
6	1	0	0	1	0	2
7	1	0	1	1	0	3
8	2	0	2	2	1	7
9	2	1	2	2	1	8
10	2	1	2	2	2	9
分散	0.178	0.767	0.678	0.178	0.622	6.844

①分散の求め方は、VARという関数を用いる。標準偏差(SD)を求めてそれを2乗しても同じである。

② α 係数を求める手順は次の通り。

問題数 = 5

問題項目の分散の合計 = $0.178+0.767+0.678+0.178+0.622=2.423$

個人得点の分散 = 6.844

α 係数 = $\frac{\text{問題数}}{\text{問題数}-1} \times (1 - \frac{\text{問題項目の分散の合計}}{\text{個人得点の分散}})$

= $5/(5-1) \times (1 - 2.423/6.844)$

= 0.807

	Q1	Q2	Q3	Q4	合計
1	2	2	1	1	6
2	2	3	2	3	10
3	1	3	1	1	6
4	3	3	2	3	11
5	1	1	3	3	8
6	2	3	2	3	10
7	2	1	3	2	8
8	3	2	3	3	11
分散	0.571	0.786	0.696	0.839	4.214
			分散の合計	2.893	
			$\alpha =$	<u>0.418</u>	

	Q1	Q2	Q3	Q4	合計
1	1	1	2	2	6
2	2	2	3	3	10
3	1	1	2	2	6
4	3	2	3	3	11
5	2	1	3	2	8
6	2	3	2	3	10
7	2	1	3	2	8
8	3	2	3	3	11
分散	0.571	0.554	0.268	0.286	4.214
			分散の合計	1.679	
			$\alpha =$	<u>0.802</u>	

内的整合法(内部一貫法)の α 係数は何を表しているのか？

★ 項目の整合性（一貫性）が高いとは何か？ どのように見るのか？

	Q1	Q2	Q3	Q4	合計		Q1	Q2	Q3	Q4	合計
1	1	2	1	1	5	1	1	1	1	2	5
2	3	1	3	3	10	2	2	2	3	3	10
3	1	3	1	1	6	3	1	1	2	2	6
4	3	2	3	3	11	4	3	2	3	3	11
5	1	3	1	2	7	5	2	1	2	2	7
6	3	1	3	3	10	6	2	3	2	3	10
7	1	3	1	1	6	7	2	1	1	2	6
8	3	2	3	3	11	8	3	2	3	3	11
分散	1.143	0.696	1.143	0.982	6.214	分散	0.571	0.554	0.696	0.286	6.214
IT相関	0.965	-0.64	0.965	0.969		IT相関	0.834	0.828	0.876	0.965	
		分散の合計			3.964			分散の合計			2.107
			$\alpha =$	0.483					$\alpha =$	0.881	

⇒ IT相関が極端に低い項目があると α 係数は低下 →項目の不整合

妥当性の検討方法

内容的 妥当性

カリキュラム 妥当性

測定で用いられている課題や質問が測定しようとする特性やカリキュラム内容を的確に反映しているかどうか

基準関連 妥当性

併存的妥当性

他の外的基準(従来のテスト, 調査, 専門家の診断など)と今回の測定との間でどの程度対応がみられるか

予測的妥当性

現在の測定結果が将来の成績や得点をどの程度予測していたか

構成概念 妥当性

因子的妥当性

測定内容の構成概念が, 因子分析結果と対応しているか

信頼性・妥当性を高めるには？

Manabu
KISHI

信頼性を高める

- ・ 解答・回答しやすいような表現, 設問
- ・ 異質な能力や特性を測定する(が必要となる) 質問項目を入れない
- ・ 出題・設問形式をなるべく揃える
- ・ 出題・設問数を増やす

妥当性を高める

- ・ 評価目的に合っているかどうかの確認
- ・ 測定の全内容(単元など)を網羅していること

妥当性の新しい考え方

「この測定の結果は大丈夫である」ことを示す証拠の
質と量によって判断 (Messickより)

- 内容的証拠 → 項目内容は構成概念に合っているか
- (本質的証拠 → 心理学的な説明は確かか)
- 構造的証拠 → 因子分析, IT相関の結果はどうか
- 一般化可能性の証拠 → 信頼性(α 係数など)はどうか
- 外的側面の証拠 → 収束的・弁別的証拠はどうか,
SEMによる検討結果はどうか
- 結果的証拠 → 予測性の確認はどうか