第１１回　古典的テスト理論（２）－測定の標準誤差・相関の希薄化

■測定の標準誤差（Standard Error of Measurement: SEM）

測定の標準誤差 ＝ 誤差の標準偏差 σE ＝ $\sigma x\sqrt{(1-\rho x)}$，$\rho x$：信頼性係数
受験者集団における標準的な誤差Ｅの大きさ
誤差が正規分布に従うとすると，
　能力値Ｔの受験者の観測得点Ｘが，Ｔ±1/2 σEに入る確率 ＝ ３８．３０％
　能力値Ｔの受験者の観測得点Ｘが，Ｔ± σEに入る確率 ＝ ６８．２６％
　能力値Ｔの受験者の観測得点Ｘが，Ｔ± ２ σEに入る確率 ＝ ９５．４４％

平均点５０，標準偏差１０，信頼性係数０．８４のテストの場合，
　測定の標準誤差 ＝ $10\sqrt{(1-0.84)} = 10 \times 0.4 = 4$
　真の得点Ｔが６０の人の観測得点が，５８～６２点に入る確率は　３８．３０％，
　５６～６４点に入る確率は　６８．２６％，
　５２～６８点に入る確率は　９５．４４％

■信頼性係数の大きさの評価の経験的な目安

　学力のような特性を考えると，例えば小学校６年生の算数でも，計算とか図形とか，測定したい特性に，ある程度の広がりがあり，計算ができても図形が苦手とか，その逆とか，各受験者において各項目の得点の一貫性を貫く（全部に正答できる，逆に，全部正答できない）ことが，もともと難しくなる．
→　信頼性係数の値は１にはならなくなる．
←　測定したい特性の広がりを確保するためには仕方ないこと．→　妥当性

信頼性から言うと理想は信頼性係数＝１
→　学力の極々一部だけを正確に測定するならそれでもよい
←　ある程度の広がりをもった学力を測定する場合には，信頼性係数は少し下げざるを得ない

経験的に言われている妥当なライン，
　個人の処遇を決めるテスト　０．９以上　（$\sigma E／\sigma x$ が ３分の１以下）
　学力・能力テスト　０．８以上　（$\sigma E／\sigma x$ が ２分の１弱）
　性格検査など　０．７以上　（$\sigma E／\sigma x$ が ２分の１強）

■相関係数の希薄化

相関の希薄化：信頼性の低い尺度を使うと，本来の相関よりも小さい相関しか観測できないこと．
X,Yの２つの変数があるとき，$rxy = \rho xy \sqrt{\rho x} \sqrt{\rho y} < \rho xy$
　rxy：観測得点間の相関係数．　ρxy：真の得点間の相関係数
　ρx：測定ｘの信頼性係数．　ρy：測定ｙの信頼性係数

複数の相関関係を見誤る可能性があるので注意
ex）　国語，社会，物理のテスト．一般的には，国語と社会の相関がまだ高い．

真の相関

	国語	社会	物理	信頼性係数
国語	1	０．８	０．６	０．７
社会		1	０．４	０．５
物理			1	０．９

観測される相関

	国語	社会	物理
国語	1	０．４７	０．４８
社会		1	０．２７
物理			1

真の相関

	国語	社会	物理	信頼性係数
国語	1	０．８	０．６	０．７
社会		1	０．４	０．７
物理			1	０．７

観測される相関

	国語	社会	物理
国語	1	０．５６	０．４２
社会		1	０．２８
物理			1

→　複数の尺度を使って研究する場合には，どの尺度の信頼性係数の大きさもなるべく同じであるのが望ましい．
　相関係数は希薄化されるが，相対的な関係性は保存される．

→　独立変数に基づいて，被験者を高群－低群と分けて平均値を比較する場合，群分け変数の信頼性が低いと，平均値差が本来よりも小さく観測されてしまう
　独立変数が複数ある場合，信頼性係数の大きい独立変数の効果が大きく観測されてしまい，どの独立変数が有効かを見誤る可能性が生じる．

希薄化の修正公式
　$\rho_{xy} = r_{xy} / (\sqrt{\rho_x \rho_y})$　を考えることはできるが，実際には使わない．
←　信頼性係数が推定値としてしか得られない．修正後の値が１を超えるようなこともある．
→　どうせやるんだったら構造方程式モデリング（Structure Equation Modeling：SEM）