



データ解析実習

2012.10.15 月曜、4・5限

担当教員：高木英至

1

今日の授業

- ▶ 1 測定
 - ▶ 1.1 測定の尺度
 - ▶ 1.2 信頼性
 - ▶ 1.3 妥当性
- ▶ 2 量的変数の基本統計量
 - ▶ 平均値から相関係数まで



▶ 2

■ 1 測定

1.1 測定の尺度

- ▶ 測定値には4種類がある

	順序	単位	原点	例
名義尺度 Nominal scale	×	×	×	職業、性別、賛成-反対
順序尺度 Ordinal scale	○	×	×	好きな順番、序数的効用、賛成-どちらでもない-反対
間隔尺度 Interval scale	○	○	×	温度、性格検査のスコア、基数的効用
比例尺度 Ratio scale	○	○	○	長さ、質量

▶ 3

- ▶ 名義尺度：値がカテゴリー
 - ▶ 例：性別、職業、国籍、民族、...
- ▶ 順序尺度：値の間に順序関係が成り立つ
 - ▶ 例：賛成-やや賛成-どちらでもない...
- ▶ 間隔尺度：値の間に順序関係があり、単位がある（値の間の間隔が定義されている）
 - ▶ 例：摂氏・華氏の温度
- ▶ 比例尺度：値の間に順序関係があり、単位があり、絶対的の原点がある
 - ▶ 例：物理的測定

▶ 4

1.2 測定値の信頼性(reliability)

- ▶ 信頼性 (reliability)
 - ▶ 同一対象に対する繰り返し測定の間に見られる測定値間の一貫性
 - ▶ 測定誤差が大きければデータの信頼性は低い
- ▶ 同じ「真の値」を測定しても、その「測定値」が一定だとは限らない
 - ▶ 例：学力テスト（測定） — その日のコンディションによって結果は変わる

5

信頼性を確認する主な方法

- ▶ (1) 再テスト法：全く同じ測定（テスト）を2度行なう
- ▶ (2) 平行テスト法：異なった、しかし同等のテストを2度行なう
- ▶ (3) 折半法(split-half method)：質問紙中の同じものを測る質問項目を2分し、その2つを平行テストと見なす
- ▶ (4) 内部一貫性による方法：質問紙中の同じものを測る質問項目間の統計量から信頼性を測定する。
 - ▶ α 係数の算出が一般的

6

信頼性係数：要点 (reliability coefficient)

- ▶ (1) $X = T + e$
 - ▶ X ：測定値、 T ：真の値、 e ：測定誤差
- ▶ (2) 指標
 - ▶ 信頼性指数 $\rho(X, T)$ ： X と T の間の相関係数
 - ▶ 信頼性係数 $\rho(X) = \sigma^2(T) / \sigma^2(X) = \rho^2(X, T)$
 - ▶ 信号－雑音比 (S-N ratio) $\rho(T/e) = \sigma^2(T) / \sigma^2(e) = \rho(X) / [1 - \rho(X)]$

7

- ▶ (3) 合成得点：複数の項目の合成（通常は加算）によって測定値とする場合
 - ▶ 一般の態度・パーソナリティの測定。
 - ▶ $X = \sum X_j$, $T = \sum T_j$, $e = \sum e_j$.
 - ▶ 同様に $\rho(X)$ を測定できる
- ▶ (3') 測定項目数 (n) が大きくなると合成得点の信頼性係数は高くなる
 - ▶ 「強平行測定」を仮定した場合、証明あり
 - ▶ 単一項目 (single item) による測定は不安定
 - ▶ 普通のパーソナリティ・テストは数多くの項目を使っている

8

- ▶ (4) 誤差を含んだ測定値間の相関係数は、真値間の相関係数より低くなる
 - ▶ 希薄化の修正公式

$$\rho(T_1, T_2) = \frac{\rho(X_1, X_2)}{\sqrt{\rho(X_1)}\sqrt{\rho(X_2)}}$$

- ▶ 誤差を含んだ測定値間の相関は、真値間の相関より低くなる

9

- ▶ (5) クロンバック (Cronbach) のα係数
 - ▶ 信頼性係数の推定値としてよく用いられる
 - ▶ 信頼性係数の一つの下限値を与える
 - ▶ $\alpha \leq \rho(X)$.
 - ▶ 内部一貫性に基づくその他の信頼性係数の推定値 — Guttman の L2 など。
- ▶ (6) 他の方法による信頼性係数の推定。
 - ▶ 再テスト法、平行テスト法による信頼性係数： 2つのテスト間の相関係数
 - ▶ 折半法による信頼性係数：α係数の求める際、測定項目数=2 としたのと同じ
 - ▶ α係数は、すべての可能な折半法の組合せによる信頼性係数の平均値になる

10

1.3 測定値の妥当性(validity)

- ▶ 妥当性 (validity)
 - ▶ 測定すべきものを実際に測定しているかどうか
 - ▶ 信頼性 — 誤差の大きさが問題
 - ▶ 信頼性が高くても妥当性が無い場合もある
- ▶ 妥当性は多義的、いろんな意味で使われる

11

様々な妥当性概念

- ▶ (1) 表面的妥当性 (face validity) : 直感的、印象的に主張される妥当性
 - ▶ [例] 特定の職務の適性検査が、当の職務内容と関連することを調べていなければ、その適性検査の表面的妥当性は低い
- ▶ (2) 内容的妥当性 (content validity) : テストの内容が測定すべき領域の適切な標本 (sample) になっていること
 - ▶ [例] 世界史の学力試験：試験問題がいろんな時代や地域を適切にカバーしていれば、試験問題の内容的妥当性は高い

12

- ▶ (3) 基準測度 (criterion measure) に基づく妥当性
 - ▶ テストの得点と基準測度との比較によって判断される妥当性
- ▶ (3-1) 予測的妥当性 (predictive validity)
 - ▶ 将来の行動の測定値が基準測度となる
 - ▶ [例] 入学試験 = 「入学後の学科履修の適性を測る」
 - ▶ 基準測度 = 入学後の成績
 - ▶ 入学試験の成績と入学後の成績の相関が高い
 - → その入学試験は予測的妥当性が高い
- ▶ (3-2) 併存的妥当性 (concurrent validity)
 - ▶ 2つのテストを併用する場合
 - ▶ [例] 精神科医の診断 (A) とパーソナリティ・テスト (B) を併用
 - ▶ 両者の相関が高ければ、A [B] は B [A] と高い併存的妥当性を持つと判断できる

13

- ▶ (4) 構成概念妥当性 (construct validity)
 - ▶ テストが予め計画された構成体 (construct) を測定していること
 - ▶ 構成体 (仮設構成体) — 直接は観測できず、研究者が「ある」と仮定したもの
 - ▶ 態度、パーソナリティ、知能、など
- ▶ (4-1) 収束的妥当性 (convergent validity)
 - ▶ 同一対象に対する測定値間の相関が高くなること
- ▶ (4-2) 弁別的妥当性 (discriminant validity)
 - ▶ 異なった対象に対する測定値間の相関が低くなること
 - ▶ 相関が高い場合、2つの測定は同じ対象を測っている可能性がある

14

- ▶ [例] X_1, X_2, X_3 — マキャベリズム的性格の測定項目
- ▶ Y_1, Y_2, Y_3 — 社会的承認欲求の測定項目
- ▶ 6つの測定項目間の相関係数の表 (相関行列) は下のようなら収束的妥当性と弁別的妥当性は高い

	X_1	X_2	X_3	Y_1	Y_2	Y_3
X_1	--					
X_2	.82	--				
X_3	.77	.63	--			
Y_1	.12	.08	-.02	--		
Y_2	.05	.18	.12	.76	--	
Y_3	.14	.21	.09	.67	.73	--

弁別的妥当性

収束的妥当性

15

- ▶ (4-3) 因子的妥当性 (factorial validity)
 - ▶ 測定項目の因子構造から判定される妥当性
 - ▶ [例] 上例の収束的妥当性と弁別的妥当性は、因子分析からも確認できる。
 - ▶ 次のような因子構造が得られれば、測定 X と測定 Y は因子的に妥当である (X は因子 1 を、 Y は因子 2 を測定していると判断できる)
 - ▶ → X と Y の測定には収束的妥当性と弁別的妥当性がある。

	因子 1	因子 2
X_1	.65	.02
X_2	.74	-.12
X_3	.68	-.08
Y_1	.01	.83
Y_2	.11	.79
Y_3	-.02	.63

16

■2 基本統計量

- ▶ 基礎統計量
 - ▶ 『心理統計学の基礎』、第1～3章
 - ▶ 『社会心理学研究入門』などに解説あり
- ▶ (1) 代表値(average)
 - ▶ 平均値 (mean)
 - ▶ メディアン
 - ▶ モード
- ▶ (2) 散布度
 - ▶ 分散・標準偏差
 - ▶ 四分位
 - ▶ 共分散
- ▶ (3) 相関係数
 - ▶ ピアソンの積率相関係数
 - ▶ 順位相関係数 (スピアマン、ケンドール...)

▶

- ▶ 平均 $\bar{X} = \frac{\sum_{i=1}^n x_i}{n}$
- ▶ 分散 $S_{xx} = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$ $S_{yy} = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-1}$
- ▶ 共分散 $S_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n-1}$
- ▶ 相関係数 $r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \cdot \sum_{i=1}^n (y_i - \bar{y})^2}} = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$

▶

共分散

$$S_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n-1}$$

- ▶ xが大きいほどyは大きい → 共分散は正の値
- ▶ xが大きいほどyは小さい → 共分散は負の値
- ▶ しかし共分散の値はx,yの尺度の取り方に依存する

▶ 19

相関係数

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \cdot \sum_{i=1}^n (y_i - \bar{y})^2}} = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$$

- ▶ 正確には「ピアソンの積率相関係数」
- ▶ 値の正負については共分散と同じ
 - ▶ xが大きいほどyは大きい → 相関係数は正の値
 - ▶ xが大きいほどyは小さい → 相関係数は負の値
 - ▶ $r_{xy}=0$ のとき、x,yは無相関(関連なし)
- ▶ しかし
 - ▶ 常に $-1 \leq r_{xy} \leq 1$
 - ▶ 相関係数の値は変数の正1次変換に対して不変

▶ 20

線形変換 (テキスト p.38)

$$x' = cx + d \quad (2.20) \quad c, d \text{ は定数}$$

$$\bar{x}' = c\bar{x} + d$$

$$s_{x'} = |c|s_x$$

$c > 0$ のとき、正 1 次変換

▶ 21

標準化：一種の線形変換

$$z = \frac{x - \bar{x}}{s_x} \quad x \text{ が素点 raw score, } z \text{ が標準得点 standard score}$$

$$z = \frac{1}{s_x} x - \frac{\bar{x}}{s_x} \quad \bar{z} = 0 \quad s_z = 1$$

$$x = \bar{x} + zs_x$$

x_i, y_i の標準得点 z_{xi}, z_{yi} をかけあわせて総和をとると、 x, y の相関係数になる。

$$\sum_{i=1}^n z_{xi} z_{yi} = \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{s_x} \cdot \frac{y_i - \bar{y}}{s_y} \right) = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} = r_{xy}$$

▶ 22

Q 1

▶ Q 1. x と y を線形変換して x' と y' を求めたとき、 x と y の相関係数と、 x' と y' の相関係数は等しくなるのはなぜか？ 説明せよ

▶ $x' = cx + d, y' = ky + l$

▶ $\rightarrow |r_{xy}| = |r_{x'y'}|$

▶ $c < 0$ の線形変換

▶ 逆転項目

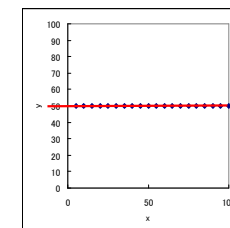
賛成	1	$\rightarrow$	5
やや賛成	2	$\rightarrow$	4
どちらでもない	3	$\rightarrow$	3
やや反対	4	$\rightarrow$	2
反対	5	$\rightarrow$	1

$$x' = 6 - x = -1 \cdot x + 6$$

▶

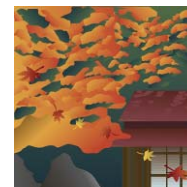
Q 2

▶ 次のような直線上にデータがあると相関係数 (r_{xy}) はどうなるのか？



▶

- ▶ Q 3. 変数 x と y があるとして、 x の値が高ければ y の値も高くなる傾向があるとき、 x と y の相関係数が正になる理由を簡単に説明しなさい。



今日は
おしまい