

# 相関分析

10/04/28

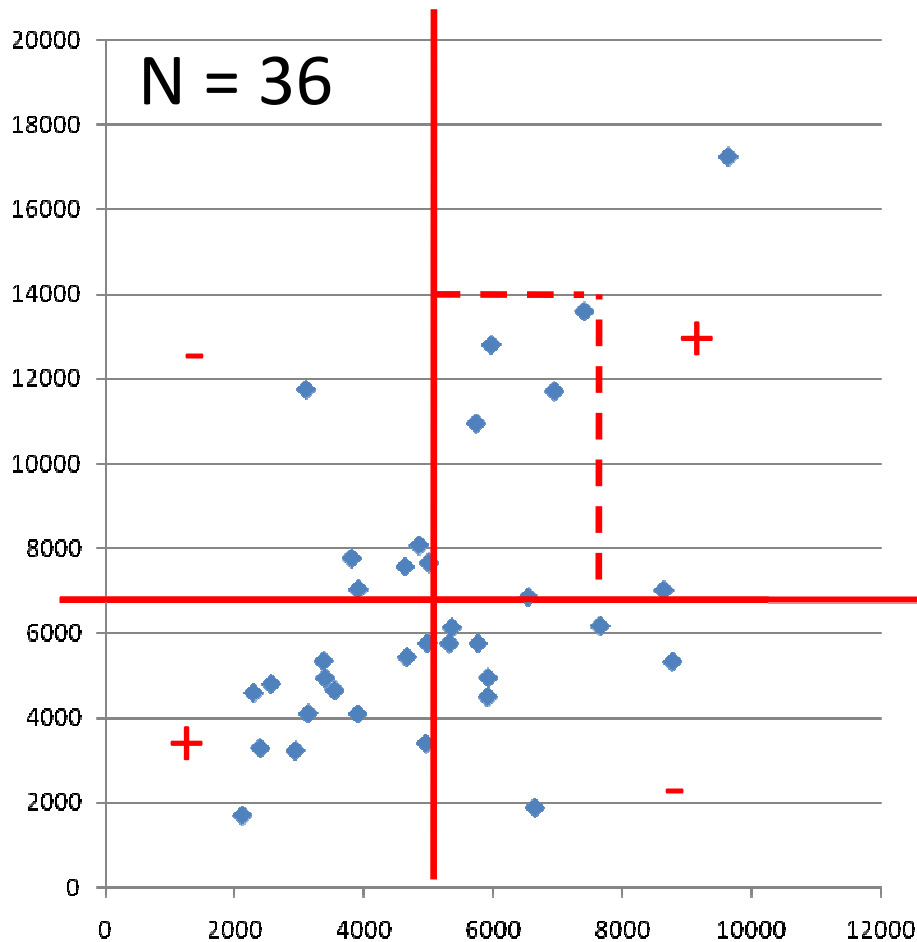
心理データ解析演習

M1 杉本匡史

# 相関分析のモデル

- 2つの変数の間の直線的な関係
  - 変数 $x$ が大き(小)くなれば変数 $y$ も大き(小)くなる +
  - 変数 $x$ が大き(小)くなれば変数 $y$ は小(大)くなる -
  - 変数 $x$ と変数 $y$ の変化に関係性がない 0
- になるような値があれば相関関係を把握しやすい

# 共分散( $s_{xy}$ )

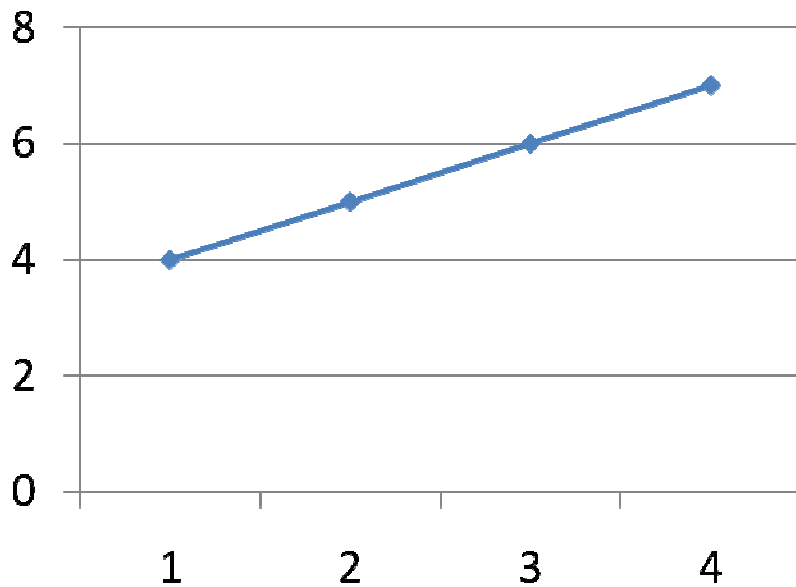


$$s_{xy} = 3290078$$

- ある1個のデータに対してそのxの値からxの平均値を引く
- さらにyの値からyの平均値を引く
- それらを掛け合わせて合計し、データ数で割る

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

- はたして3290078という共分散は大きいのか？
- 共分散が取りうる最大値と比較してみる
- では共分散が最大値, つまり2つの変数の間に完全な相関がある時の共分散は？



- 完全な対応がある時  
 $y = ax + b$ と表せる

$$\begin{aligned}
 s_{xy} &= \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y}) \\
 &= \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x}) \{ (cx_i + d) - (c\bar{x} + d) \} \\
 &= \frac{c}{N} \sum_{i=1}^N (x_i - \bar{x})^2 \\
 &= cs_x^2 \\
 &= (s_y / s_x) s_x^2 \\
 &= s_x s_y
 \end{aligned}$$

# 相関係数のモデルまとめ

- 共分散が取りうる最大値に対して実際の共分散がどの程度になっているのか

相関係数

$$r = \frac{S_{xy}}{S_x S_y}$$

# 相関係数の有意性検定

- 母集団における相関係数  $\rho$  が0かどうかの検定
- 帰無仮説:  $\rho = 0$
- 対立仮説:  $\rho \neq 0$  (場合によっては  $\rho < 0$ )
- 下の式が自由度  $N-2$  の  $t$  分布に従うという性質を利用

$$t = \frac{r}{\sqrt{\frac{1-r^2}{N-2}}}$$

この部分のために、 $N$ が大きくなればなるほど  $r$  が小さくても有意になってしまう  
cf.  $N=1000$  のとき  $r=.062$  でも有意

# 統計的に有意な相関

- 統計的に有意な相関とは？
  - が0と等しくない
- しかしNの値によって $p < .05$ となる $r$ の値は大きく変化する
  - $p$ 値だけで相関を判断するのはまずい
- ではどうすれば2つの変数の関係の強さを正確に評価できるのか？



# 心理学的に意味のある相関

- 分散説明率:  $r^2$
- 従属変数の分散の内, 独立変数の分散で説明できる部分は  $r^2/1$ , 残差で説明できる部分が  $(1 - r^2)/1$
- $N=1000$  で  $r = .062$  だけど  $p < .05$  になった
- $r^2 = (.062)^2 = .004$  従属変数の分散のうち 0.4% を独立変数で説明できた...

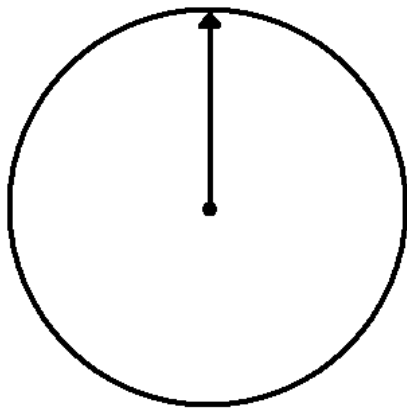
# 妥当性

- 測定値が測定対象を正確に測定できているかどうか
- 妥当性が低い場合, 相関係数が低くなる場合と高くなる場合がある
- 低くなるケース

- 相関係数が不当に高くなってしまうケース
- 2つの変数が同じ攪乱要因を共有
- 回答の際に対象としている能力とは無関係な技能を要する手法を用いる場合
- 例:日本人の国語能力と数学能力を測定するためにテストを実施するが、そのテストが英語で表記されている
- 2つの課題の形式が似通っている場合
- Fields & Shelton (2006)

## 変数x(視点取得能力)の測定方法

“imagine you are standing near the stop sign facing the house; point to the traffic light”



## 変数y(空間表象構築能力)の測定方法

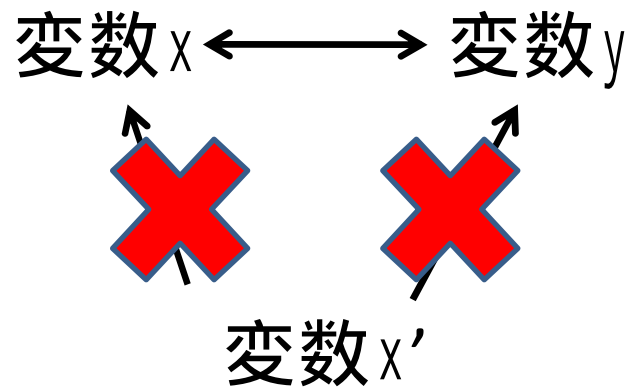
Imagine you are at the carousel and facing the jungle gym.

Point to the fountain, then click the mouse button.



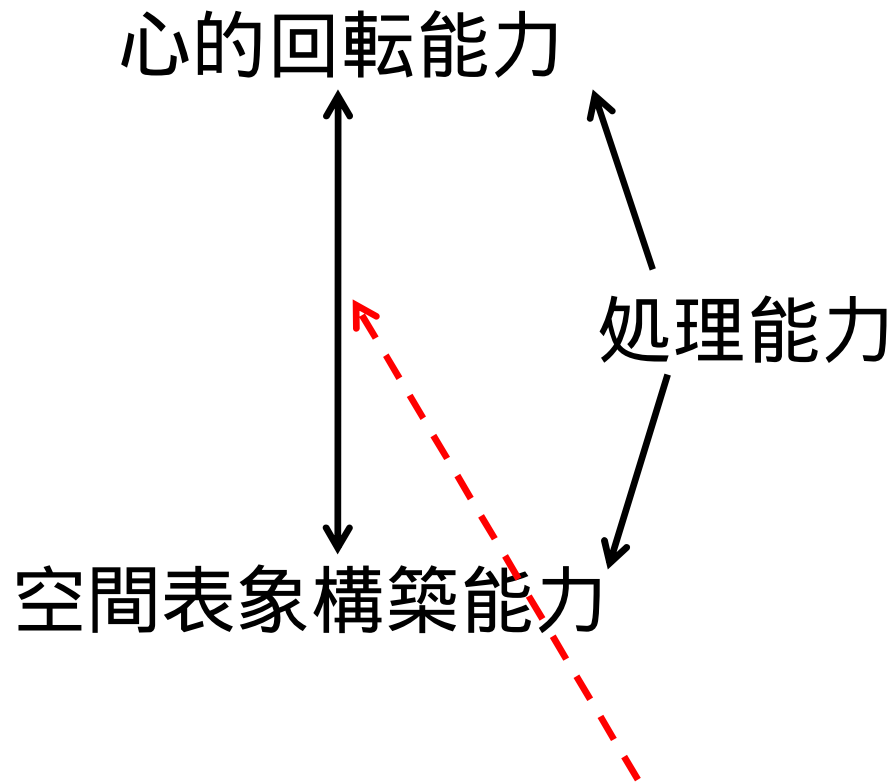
Figure 3. The judgments of relative direction task.

# 偏相関



- 変数 $x$ と変数 $y$ の間を関係を見たいのに変数 $x'$ が $x$ と $y$ に影響してしまう
- 疑似相関の可能性
- 偏相関を使って対象としていない変数の影響を取り除いてしまう
- そうすれば $x$  $y$ 間の純粋な相関がみれるはず

# 実例



- 空間についての文章を読んで空間表象を構築してもらい、その能力と心的回転の相関を調べたい
- しかし正誤問題に対する一般的な処理能力の高低が2つの変数に影響を与えている可能性は？

知りたいのはここから処理能力による影響分を除いた相関

- 空間情報を含まない正誤問題(非空間問題)の成績を一般的な処理能力として統制してみる
- 分析 相関 二変量
  - 心的回転・空間・非空間を変数に投入してOK
- 分析 相関 偏相関
  - 心的回転・空間を変数に, 非空間を制御変数に投入してOK

\*無題2 [データセット1] - SPSS データ エディタ(D)

ファイル(F) 編集(E) 表示(V) データ(D) 変換(T) 分析(A) グラフ(G) ユーティリティ(U) ウィンドウ(W) ヘルプ(H)

1: 性別 m

	性別	参加者番号	心的回転
1	m	1	7654.3
2	m	2	5001.1
3	m	3	2283.6
4	m	4	2929.5
5	m	5	4632.0
6	m	6	3900.9
7	m	7	6388.6
8	m	8	6948.1
9	m	9	3130.0
10	m	10	5921.7
11	m	11	4976.3
12	m	12	8774.3
13	...	13	5919.7

分析(A) のメニュー:

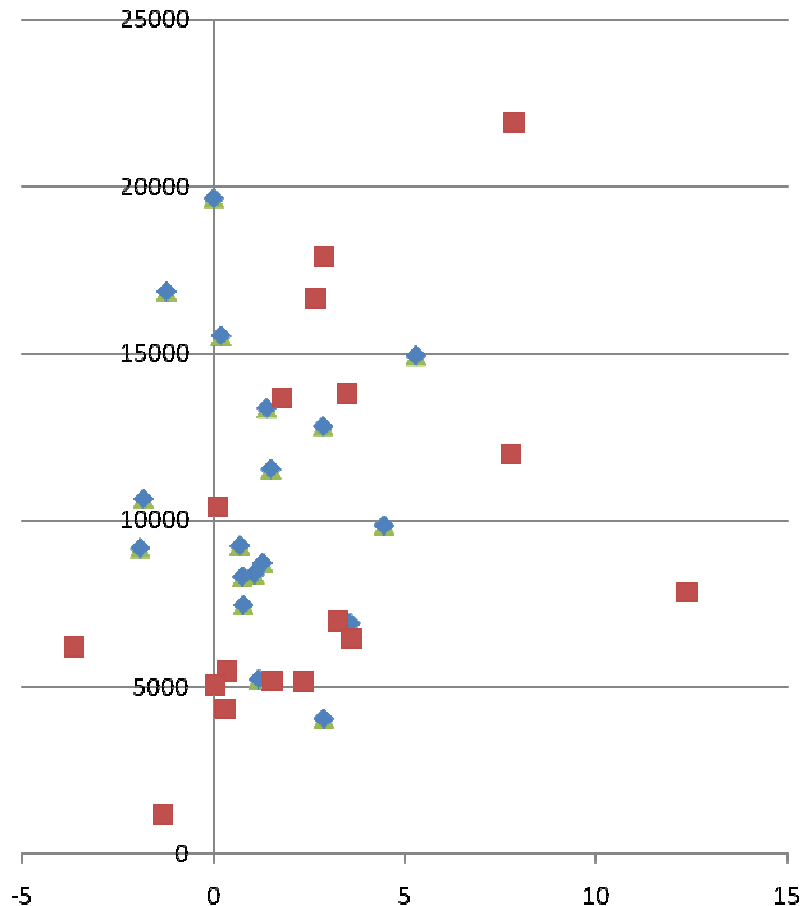
- 報告書(P)
- 記述統計(E)
- テーブル(B)
- 平均の比較(M)
- 一般線型モデル(G)
- 一般化線型モデル(Z)
- 混合モデル(X)
- 相関(C)**
  - $r_{12}$  2変数(B)...
  - $r_{12,3}$  偏相関(R)...
  - $\delta$  距離(D)...
- 回帰(R)
- 対数線型(O)
- 分類(F)
- データの分解(D)
- スケール(A)



# 偏相関：結果

- $r = .500$ から $r = .388$ に低下
- $r^2$ も同様に.250から.151に低下
- 何が言える？

# 分割相関



縦軸：空間表象構築能力  
横軸：視点取得能力

- 何らかの第3の変数に関する群ごとに $r$ を算出すると、群ごとの $r$ が異なることがある
- 全体での $r = .278$
- 男性参加者の $r = -.169$
- 女性参加者の $r = .450$

# 2つの相関係数の差(異なる標本に基づく場合)

- 相関係数をフィッシャーの $z'$ 変換して  $\chi^2$ 検定

$$z'_i = \frac{1}{2} \log e \frac{1+r}{1-r}$$

- 下記HPでも可能

– [http://www.hs.hirosaki-u.ac.jp/~pteiki/research/stat/cor\\_test.xls](http://www.hs.hirosaki-u.ac.jp/~pteiki/research/stat/cor_test.xls)

- $\chi^2 = z'^2(N-3) - \{z'(N-3)\}^2/N-3$

# 実際のデータ

	r	z'	N-3	z'(N-3)	z' <sup>2</sup> (N-3)
男性	-.169	-.171	15	-2.565	0.439
女性	.450	.485	14	6.79	3.293
合計			29	4.225	3.732

- $df = 1$  のとき  $\chi^2$  の臨界値は
  - 10%水準で2.71
  - 5%水準で3.84
  - 1%水準で6.63

## 相関の差(異なる標本の場合): 結果

- $s^2=3.732 - (4.225)^2/29=3.1164$
- 10%水準で有意な相関の差がみられた
- 今回のケースではNが小さいために実用的ではない可能性がある

## 2つの相関係数の差(同じ標本に基づ く場合)

- $r_{12}$ と  $r_{13}$ において

$$t = (r_{12} - r_{13}) \sqrt{\frac{(n+3)(1+r_{23})}{2(1-r_{12}^2 - r_{13}^2 - r_{23}^2 + 2r_{12}r_{13}r_{23})}}$$

- $df = N-3$

# 実際のデータ

- 実際のデータ (N = 35)

	心的回転	問題A	問題B
心的回転	1	0.500	0.197
問題A		1	.546
問題B			1

- $df = 32$  のとき  $t$  の臨界値は
  - 10%水準で1.69
  - 5%水準で2.04
  - 1%水準で2.74

# 相関の差(同じ標本の場合)

- $t=2.2759$
- 5%水準で有意な相関の差がみられた
- ということは？



# 参考文献

- 岩原信九郎 (1957) 教育と心理のための推計学 日本文化科学社
- 南風原朝和 (2002) 心理統計学の基礎 統合的理解のために 有斐閣アルマ
- 村瀬洋一・高田洋・廣瀬毅士 (2007) SPSSによる多変量解析 オーム社
- 森敏昭・吉田寿夫 (1990) 心理学のためのデータ解析 テクニカルブック 北大路書房
- SPSSで学ぶ医療系データ解析(弘前大学大学院保健学研究科 津島研究室) <http://www.hs.hirosaki-u.ac.jp/~pteiki/research/stat/text.html>