

質問票開発のための統計学2

国立がん研究センター 研究支援センター 生物統計部
野村 尚吾

復習：質問票に求められる性質

■ 妥当性 (validity)

- 本当に測定したいものを測定しているか？
- ただし、ほぼすべての状況で絶対的な真値は存在しない

■ 信頼性 (reliability)

- 測定値のバラツキの評価
- 同じ結果が再現されるかどうか？

■ 感度 (sensitivity)と反応性 (responsiveness)

- 感度：集団間での違いを検出できるか？
(例：一般集団とがん患者でQOLスコアの分布が分かれるか？)
- 反応性：ある患者の病態推移が捉えられるか？
(例：病状の悪化に伴ってQOLスコアが経時的に変化するか？)

復習：妥当性を評価する方法

■ 正解がある場合

- **基準関連妥当性 (criterion validity) の評価**
 - 既に確立された質問票や実際に観察される症状等との一致度を評価する
- よく用いられる統計手法
 - 相関係数
(一致の強弱を評価できない点に注意が必要)
 - 診断の場面で用いられる手法 (感度・特異度など)

■ 正解がない場合

- **構成概念妥当性 (construct validity) の評価**
- よく用いられる統計手法
 - **相関係数、Cronbachの α 係数、因子分析**

復習：信頼性を評価する方法

- 正解の有無によって評価方法は大きく変わらない
- 評価者内信頼性（再テスト信頼性）
 - 同じ対象者を経時的に評価して1回目・2回目の結果を比較
- 評価者間信頼性
 - 同じ対象者を異なる評価者が評価して結果を比較
- よく用いられる統計手法
 - スコアが連続変数の場合： 級内相関係数（ICC）
 - スコアがカテゴリー変数の場合： カッパ係数

■ 質問票開発で用いられる統計手法

- － 妥当性評価
 - 相関係数
 - Cronbachの α 係数
- － 信頼性評価
 - 級内相関係数
 - カッパ係数

まず、相関係数について

■ ピークフローデータ

– ピークフロー：呼吸量（吐く息のみ）の測定機器

■ 7名のぜんそく患者に2回繰り返し測定したデータ

（注：SKETCH研究会統計分科会(2005) p.4より引用）

患者	1	2	3	4	5	6	7
1回目	280	540	400	490	520	220	350
2回目	310	440	430	420	450	340	390

単位：L/分



Pearsonの相関係数

■ 平方和 = (平均からのズレ)² の総和

患者	1	2	3	4	5	6	7	平均
1回目(x)	280	540	400	490	520	220	350	<u>400</u>
2回目(y)	310	440	430	420	450	340	390	<u>397.1</u>

– 1回目測定 of 平方和 (S_{xx}) = $(280 - 400)^2 + (540 - 400)^2 + \dots + (350 - 400)^2 = 91400$

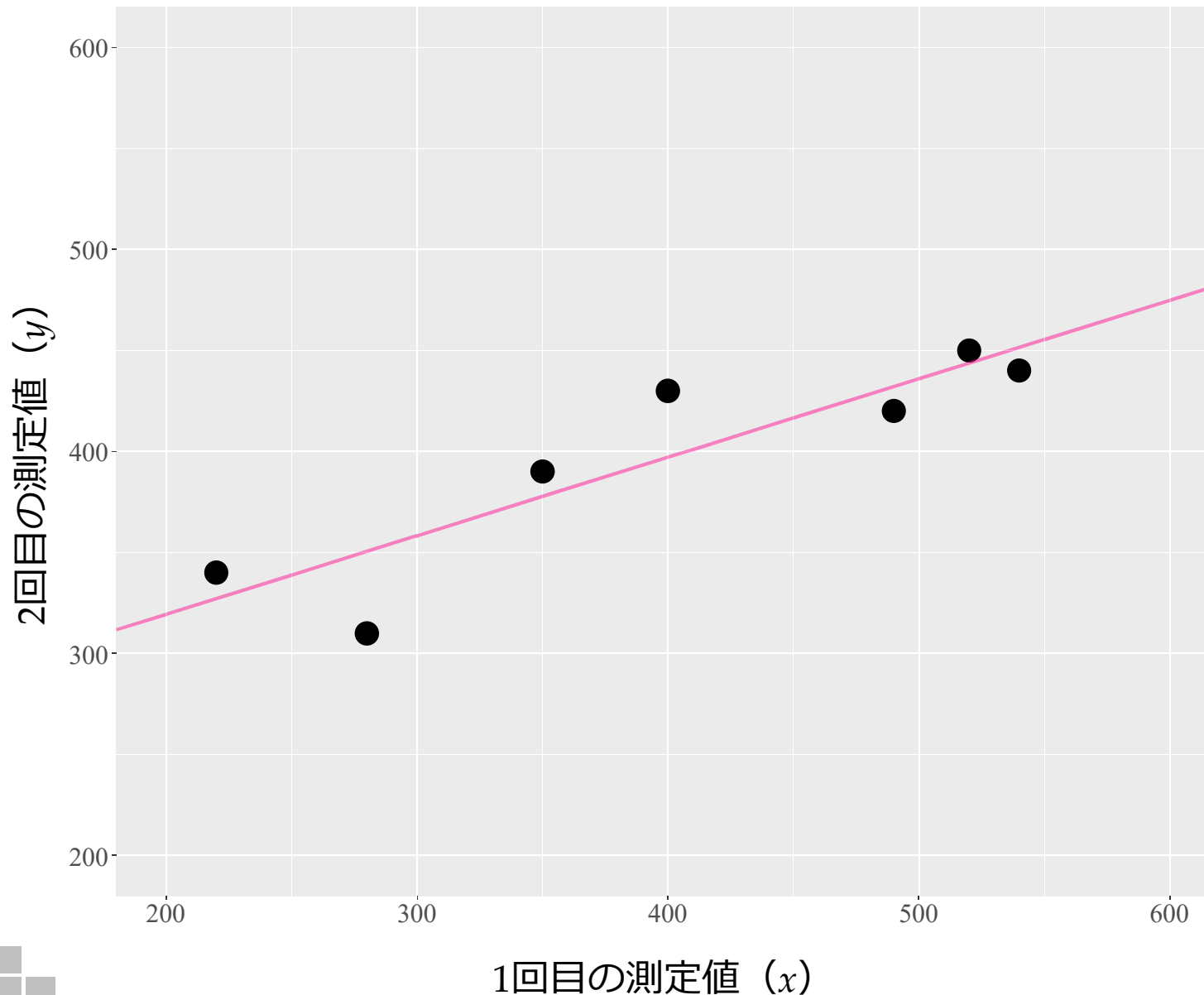
– 2回目測定 of 平方和 (S_{yy}) = 17143

■ 関連の強さを調べるために計算するもの：偏差積和 (S_{xy})

$$(280 - 400) \times (310 - 397.1) + (540 - 400) \times (440 - 397.1) + \dots + (350 - 400) \times (390 - 397.1) = 35500$$

■ Pearsonの相関係数：
$$\rho = \frac{S_{xy}}{\sqrt{S_{xx} S_{yy}}} = 0.897$$

回帰直線と相関係数の関係



$$y = 241.8 + 0.388x$$

$$\text{傾き} = \frac{S_{xy}}{S_{xx}} = \frac{35500}{91400}$$

$$\text{寄与率} R^2 = \frac{S_{xy}^2}{S_{xx} S_{yy}} = 0.897^2$$

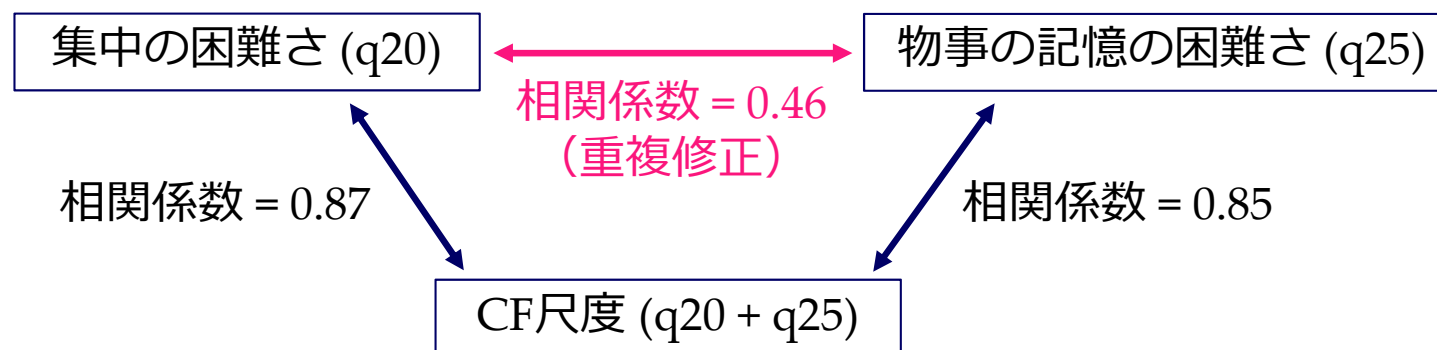
相関係数：妥当性評価での注意点 1

- **平均値による要約が不適切なとき（カテゴリー尺度など）
Pearsonの相関係数は2変数間の関連を適切に反映しない**
 - 対数変換などの変換を行ってからPearsonの相関係数を使う
 - （順位情報に基づく）Spearmanの相関係数やKendallの相関係数を使う
- **十分なサンプルサイズで調査されていない場合は
相関係数の推定精度も評価する必要がある**
 - 福原(2005)では100例以上を1つの目安としている（p.78）

相関係数：妥当性評価での注意点 2

■ 重複の修正

－ 例：EORTC QLQ-C30 CF（認知機能尺度）



合計スコアと各項目の相関を評価する場合に重複の修正が必要

相関係数：妥当性評価での注意点 3

- 「ある値（例：0.6）より高いと相関が高い」という絶対的な評価に向かない → **相対的な評価が重要**
 - 仮想例：頭頸部癌におけるQOLスコアの比較
 - 既存のQOL尺度 "A" と新規のQOL尺度 "B" の比較（いずれも連続量とする）
 - 複数部位で基準関連妥当性を調べた結果（Pearsonの相関係数）
 - 集団1（咽頭がん, n=100）： $\rho=0.48$
 - 集団2（喉頭がん, n=100）： $\rho=0.51$
 - 集団3（それ以外, n=100）： $\rho=0.44$
 - 集団1～集団3（n=300）： $\rho=0.93$
- 相関係数が低く、妥当性が低い
- 相関係数が高いため
十分な妥当性が認められる？

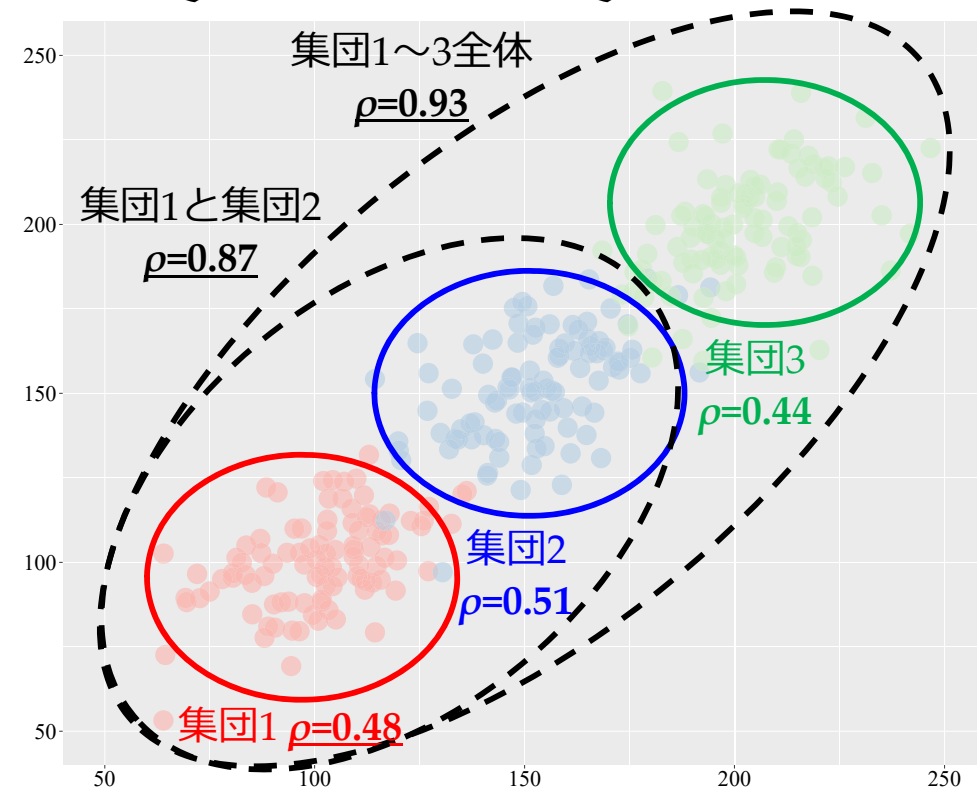
■ 同質集団よりも異質集団の方が相関係数が高くなりやすい

- 「統一された疾患状態をもった同質集団」よりも、
「様々な疾患の状態をもった異質集団」の方が範囲が広がる

－ 例：2変量正規分布

- 集団1：平均=100, SD=50, $\rho=0.48$
- 集団2：平均=150, SD=50, $\rho=0.51$
- 集団3：平均=200, SD=50, $\rho=0.44$
- 集団1と集団2： $\rho=0.87$
- 集団1～集団3： $\rho=0.93$

縦軸：QOL尺度A、横軸：QOL尺度B



※ 各サブ集団のサンプルサイズは100

- － 特定集団内で相関係数の相対的な大きさを評価すべき

相関係数：妥当性評価での注意点 4

例：SF-36の収束的・弁別的妥当性 (n=3445)

■ 構成概念妥当性：相関マトリックスで評価 (福原(2005), p.77)

表 4.1 重複による修正を適用したときの各項目と仮定される尺度との相関係数

項目	尺度の相関係数								
	PF	RP	BP	GH	VT	SF	RE	MH	TRAN
RP1	0.58	0.67	0.55	0.42	0.43	0.41	0.28	0.20	0.12
RP2	0.51	0.70	0.51	0.42	0.47	0.46	0.40	0.27	0.17
RP3	0.44	0.65	0.46	0.39	0.51	0.39	0.41	0.28	0.12
RP4	0.49	0.68	0.56	0.41	0.50	0.45	0.39	0.32	0.14
BP1	0.53	0.54	0.70	0.44	0.46	0.43	0.27	0.29	0.21
BP2	0.53	0.61	0.70	0.44	0.48	0.51	0.33	0.33	0.19

PF：physical functioning/身体機能，RP：role physical/日常役割機能(身体)，BP：bodily pain/体の痛み，GH：general health perceptions/全体的健康感，VT：vitality/活力，SF：social functioning/社会生活機能，RE：role emotional/日常役割機能(精神)，MH：mental health/心の健康，TRAN：reported change in health/健康の推移。

McHorney ら (1994) による

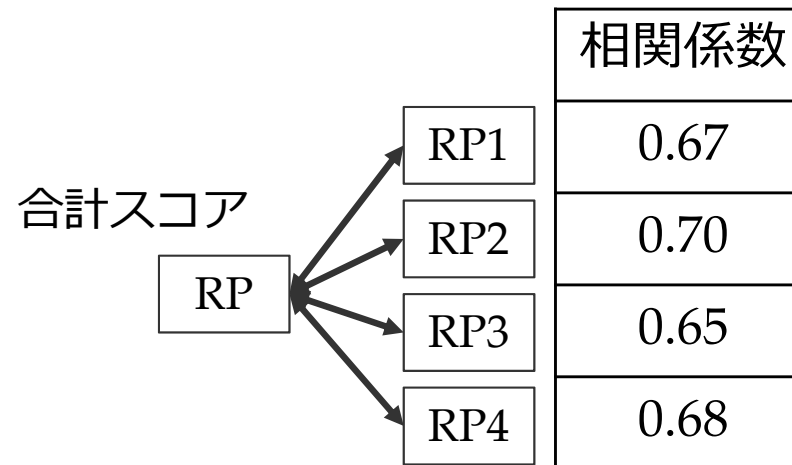
重複修正済み相関係数が高い
(RPドメイン内の相関が高い)
→ 収束的妥当性

他ドメインとの相関が低い
→ 弁別的妥当性

相関係数で分からないこと

例：SF-36の収束的・弁別的妥当性 (n=3445)

■ 例：SF-36の重複修正済みの相関係数（RPドメイン）

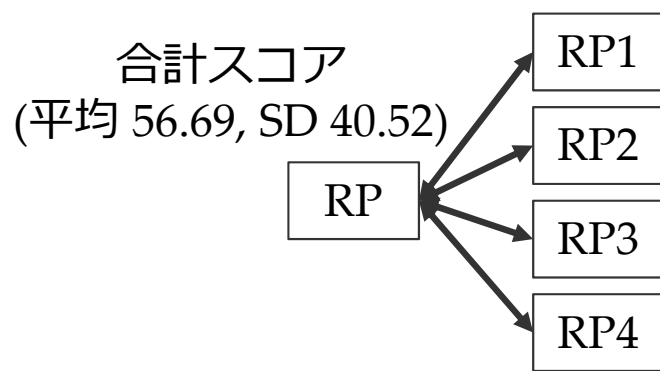


■ プロファイル型質問票では合計スコアが一般的だが、そもそも合計スコアは妥当なのか？

- 相関係数が高い値を示したからと言って、RP1～4が同じ「日常活動機能」を測定している保証はない
- 「RP1が1単位増えた場合」と「RP2が1単位増えた場合」でRPの変動が異なるなら、合計スコアは合成尺度として妥当でない

■ 内的整合性の評価

- 内的整合性 = 各項目が同じドメインを測定していること
 - これは（重複修正済みの）相関係数だけでは分からない
- まず、平均・SD・欠測割合等のデータの要約情報を参照する
 - 分布の形が似ていれば内的整合性が担保される可能性は高い



	平均	SD	相関係数	欠測(%)
RP1	1.63	0.48	0.67	4.4
RP2	1.62	0.49	0.70	4.3
RP3	1.46	0.50	0.65	4.8
RP4	1.57	0.50	0.68	4.0

- それ以外の方法として、**Cronbachの α 係数**が頻用される

内的整合性の評価：イメージ

■ 例：RP（SF-36の日常役割機能評価）

患者 i のRP1 = 患者 i の日常活動機能 + 誤差

患者 i のRP2 = 患者 i の日常活動機能 + 誤差

患者 i のRP3 = 患者 i の日常活動機能 + 誤差

患者 i のRP4 = 患者 i の日常活動機能 + 誤差

真実の値

- RP1～4が日常活動機能の真実の値を測定しているなら、上の式における「誤差」はすべて小さな値になるはず
- RP1～4の相関係数
 - 誤差がゼロなら、1
 - 各項目の大部分を誤差が占めるとき0
 - これを定量化した指標がCronbachの α 係数

Cronbachの α 係数

■ 内的整合性を評価するための要約指標

$$\alpha = \frac{p}{p-1} \times \frac{\sum_{j=1}^p \sum_{k \neq j}^p \rho_{jk}}{p + \sum_{j=1}^p \sum_{k \neq j}^p \rho_{jk}}$$

p : 項目数

ρ_{jk} : 項目 j と項目 k のPearson相関係数

■ 解釈

- α は0から1の値をとり、値が大きいと内的整合性が高い
- $\alpha = 0$: 対象者 i のスコアがすべてランダムな誤差
- $\alpha = 1$: 対象者 i の全項目で誤差がゼロ

例：SF-36の収束的・弁別的妥当性 (n=3445)

表 4.2 項目尺度化に関する検定：SF-36 尺度に対する収束的妥当性および弁別的妥当性

項目	項目数	収束的妥当性 (相関係数の範囲)	弁別的妥当性 (相関係数の範囲)	尺度化の成功	尺度化成功率	等質性 (平均項目間相関)	信頼性 (Cronbach α)
PF	10	0.49-0.80	0.10-0.54	80/80	100	0.56	0.93
RP	4	0.65-0.70	0.12-0.58	32/32	100	0.57	0.84
BP	2	0.70	0.19-0.61	16/16	100	0.70	0.82
GH	5	0.38-0.72	0.09-0.58	40/40	100	0.42	0.78
VT	4	0.69-0.75	0.17-0.55	32/32	100	0.62	0.87
SF	2	0.74	0.20-0.62	16/16	100	0.74	0.85
RE	3	0.63-0.73	0.11-0.56	24/24	100	0.61	0.83
MH	5	0.65-0.81	0.11-0.59	40/40	100	0.64	0.90

RPの構成概念妥当性を支持するデータ

- RPドメイン内の相関係数が他ドメイン内の相関より高い（収束的妥当性と弁別的妥当性）
- RPドメイン内の各項目の平均・SD・（重複修正済み）相関係数は似た値を示している
- RPドメインのCronbachの α 係数の値は十分高い値

同様の傾向はSF-36の全ドメインについても確認された

小括：構成概念妥当性評価のための統計手法

■ 相関マトリックスのチェック

- ただし、相関係数の利用・解釈には注意が必要
 - 分布型にあった相関係数が選択されているか？
 - 相関係数の推定精度も検討に入れたか？
 - 重複の修正が必要か？
 - 均一な集団で相関係数が計算されているか？

■ 内的整合性のチェック

- Cronbachの α は内的整合性を評価する1つの指標に過ぎない
- 平均[中央値]・SD[Range]・重複修正済み相関係数も含めた総合的な評価が必要

■ 割愛したが、因子分析も重要

■ 質問票開発で用いられる統計手法

- － 妥当性評価
 - 相関係数
 - Cronbachの α 係数
- － 信頼性評価
 - 級内相関係数
 - カッパ係数

質問票開発のための統計手法

■ 妥当性評価のための手法

- 相関係数
- Cronbachの α 係数
- 因子分析（今回は割愛）

■ 信頼性評価のための手法

- 級内相関係数（Intraclass correlation coefficient ; ICC）
- カッパ(κ)係数

再掲：ピークフローデータ

■ ピークフロー

- ぜんそく患者の呼吸量（吐く息のみ）を測定する機器

■ 7名のぜんそく患者に2回繰り返し測定したデータ

（注：SKETCH研究会統計分科会(2005) p.4より引用）

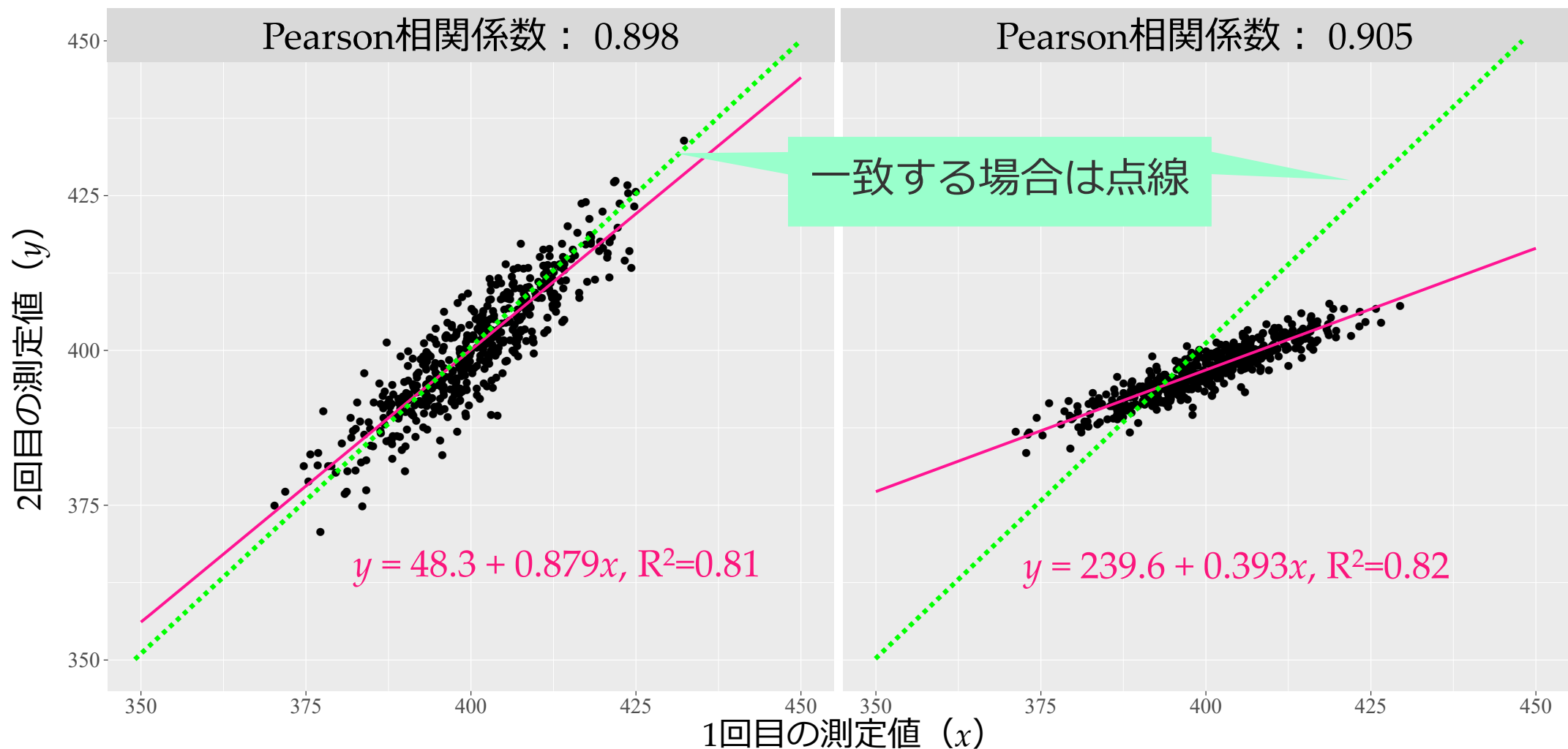
患者	1	2	3	4	5	6	7
1回目	280	540	400	490	520	220	350
2回目	310	440	430	420	450	340	390

単位：L/分

別コホートで500名を集めて評価者内信頼性を調べたところ、
高い相関関係が示されたとする・・・



■ 相関係数 = 0.9が観察される2パターンのデータ (n=500)



相関係数は再現性を評価する指標ではない

信頼性評価で重要なこと

■ 繰り返して測定した値が一致すること

- 前ページの緑点線(45°の対角線)にデータが並ぶのが理想
- これは相関係数では評価できない

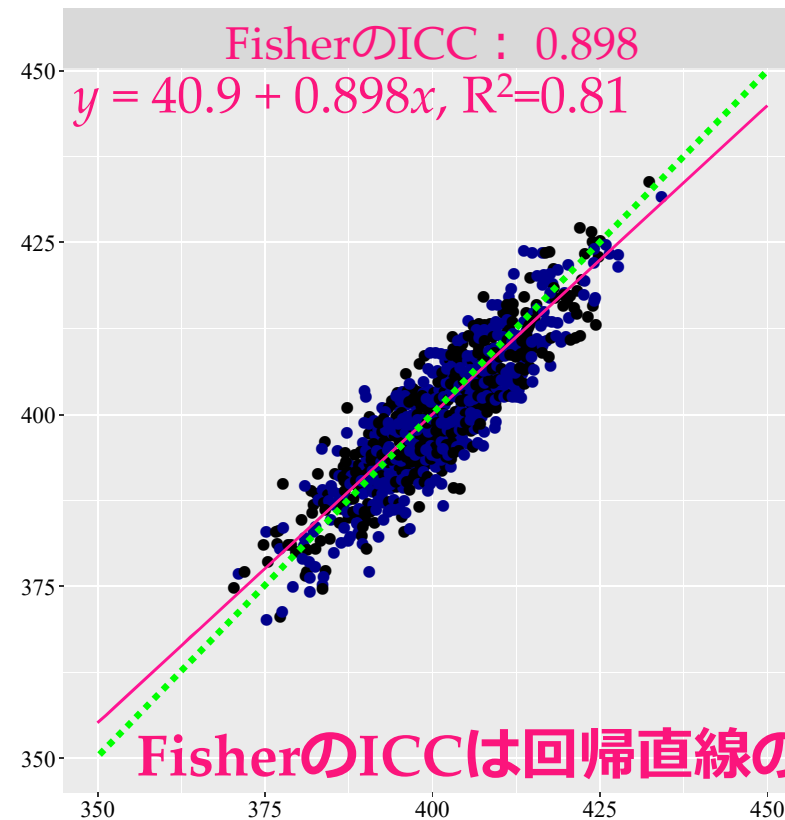
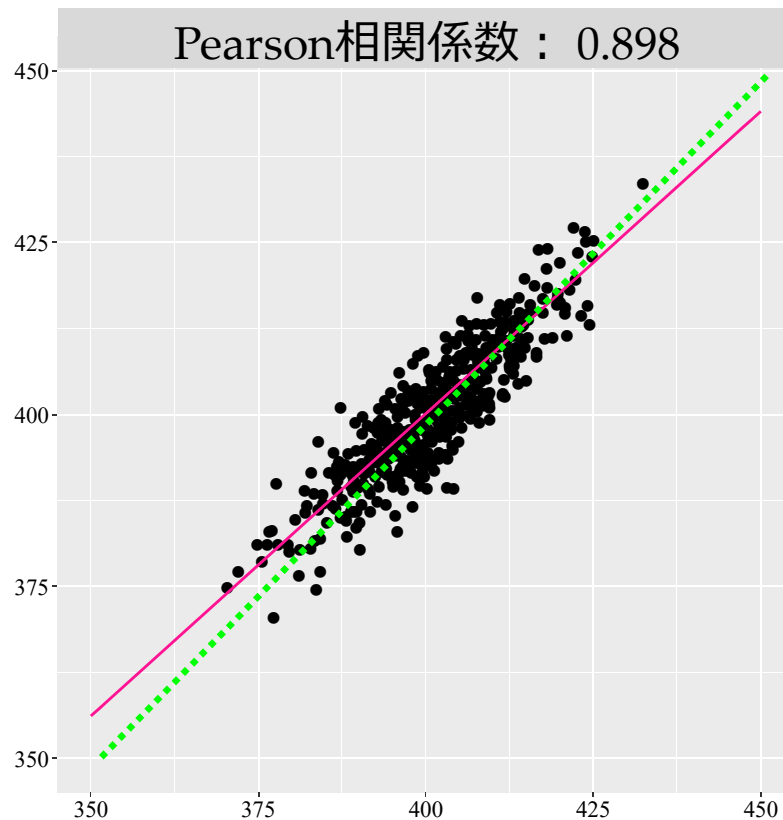
■ 「何回目か？」の情報に左右されない一致度が知りたい

- **級内相関係数 (intra-class correlation coefficient; ICC)**が頻用される

注：被験者間信頼性の場合は「評価者が誰か？」の情報に左右されない一致度を評価する（同じ考え方が使える）

級内相関係数 (ICC)の考え方

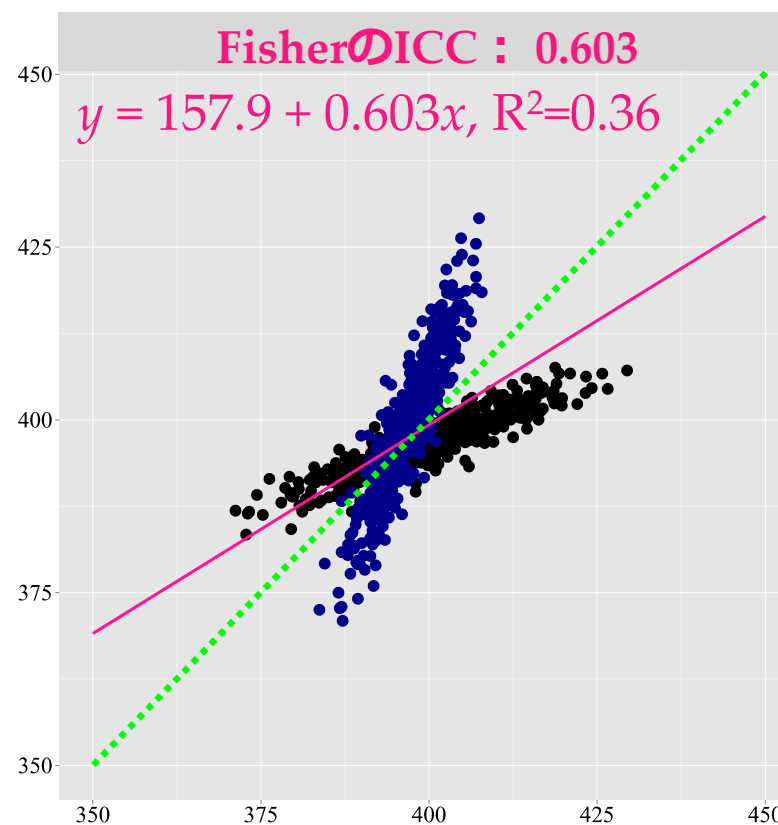
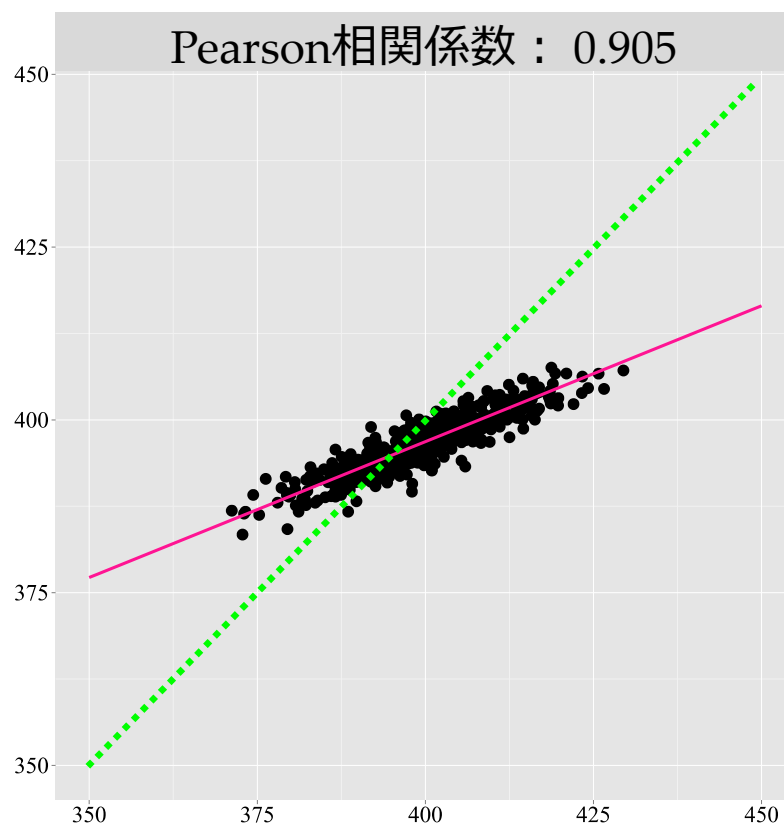
- **対角線（緑点線）に近づくほど値が大きくなる相関係数**
 - 対角線を(線)対称に移動したデータも含めてPearson相関係数を計算
 - 左図：先ほどの500例のデータ、右図：1000例のデータ



FisherのICCは回帰直線の傾きと一致

■ 相関係数は高いが一致度が低いケース

- 左図：元々のデータセット ($\rho = 0.905$)
- 右図：線対称にしたデータも含めたデータセット ($\rho = 0.603$)



緑点線にデータが並んでいない場合、ICCの値は小さくなる

ICCに関する補足事項

- 教科書には、
「ICCは対象の効果を固定効果/変動効果とした分散分析を用いて計算する」といった難しい説明が掲載
- 「**ICC = 線対称に移動したデータを含めた相関係数**」
という理解で問題ない
 - **FisherのICCと(一元配置)分散分析で計算したICCはサンプルサイズが十分に大きいとき一致する関係にある**
- **ICCは状況に応じて様々な算出法があります。
実際に算出される際には生物統計家に相談下さい。**

カテゴリー変数の信頼性評価：カッパ係数

注：仮想例

■ 画像診断データ

- 30人の画像データを2回繰り返し診断したデータ
- 対角線が一致（**観測一致割合：21/30 = 70%**）

		2回目					合計
		診断1	診断2	診断3	診断4	診断5	
1回目	診断1	4	1				5
	診断2	2	3	1			6
	診断3	1	1	4	1		7
	診断4			1	6	1	8
	診断5					4	4
合計		7	5	6	7	5	30

一致割合の問題点

- 適当に診断しても偶然に一致することがある
 - － 「偶然に一致する割合」と対比して評価する必要がある

		2回目					選択割合
		診断1	診断2	診断3	診断4	診断5	
1回目	診断1	一致					20%
	診断2		一致				20%
	診断3			一致			20%
	診断4				一致		20%
	診断5					一致	20%
合計		20%	20%	20%	20%	20%	30

- － 上の表で偶然に一致する割合（期待一致割合）
 - $20\% \times 20\% = 4\%$
 - 30人中、1.2人は平均的に一致することになる

画像診断データの期待一致割合

		2回目					合計 (%)
		診断1	診断2	診断3	診断4	診断5	
1回目	診断1	4	1				5 (17%)
	診断2	2	3	1			6 (20%)
	診断3	1	1	4	1		7 (23%)
	診断4			1	6	1	8 (27%)
	診断5					4	4 (13%)
合計 (%)		7 (23%)	5 (17%)	6 (20%)	7 (23%)	5 (17%)	30

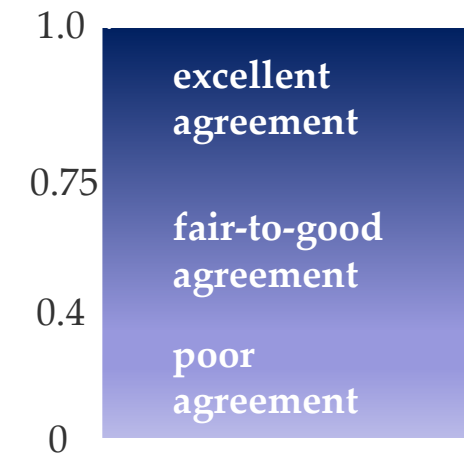
期待一致割合 $= 0.1667 \times 0.2333 + 0.2 \times 0.1667 + 0.2333 \times 0.2 + 0.2667 \times 0.2333 + 0.1333 \times 0.1667$
 $= 0.2033$

カッパ係数

$$\kappa = \frac{\text{観測一致割合} - \text{期待一致割合}}{1 - \text{期待一致割合}} = \frac{0.7 - 0.2033}{1 - 0.2033} = 0.6233$$

- 偶然一致する割合を除いた上での一致度
- 参考：一致度の強さ分類（右図参照）
- (単純)カッパ係数の問題点
 - 多分類だと一致する可能性が低い
 - 多分類の場合は1段階・2段階のズレを考慮した**重み付きカッパ係数**を用いるのが一般的

Landis & Kochの分類



まとめ

- 妥当性評価に用いる統計手法
 - 基本は相関係数 or 感度/特異度など
 - 内的整合性の評価：Cronbachの α やその他の要約指標(SD等)
- 信頼性評価に用いる統計手法
 - 級内相関係数
 - カッパ係数

本日触れられなかった統計手法

■ 因子分析

■ 項目応答理論

- 興味のある方は以下の文献を参照
 - 福原(2005), 6章
 - SKETCH研究会統計分科会(2005), 10章

■ 欠測データの解析手法

- 発展編の講義「臨床研究における欠測データの統計解析」を参照

<https://www.icrweb.jp/course/view.php?id=309>

※本日の講義で取りあげた主なトピックは以下の書籍でより詳しく学ぶことができます



大橋靖雄, 2011
医歯薬出版
ASIN: B00I399SY4



SKETCH研究会統計
分科会, 2005
サイエンス社
ISBN-13: 978-4860790226

- 正確度と精密度 p.3-8
- 信頼性と妥当性 p.24-32

- 級内相関係数 p.1-33
- カッパ係数 p.34-66

※生物統計セミナー【入門編】をより深く理解していただくための推薦図書



浜田知久馬. 2012
真興交易医書出版部; 新版
ISBN-13: 978-4880038612



佐藤俊哉. 2005
岩波書店
ISBN-13: 978-4000074544



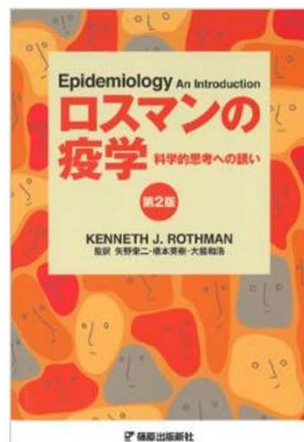
佐藤俊哉. 2012
岩波書店
ISBN-13: 978-4000295949



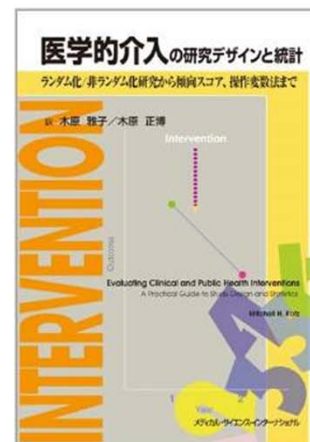
福田治彦 (訳). 2013
医学書院; 第2版
ISBN-13: 978-4260018647



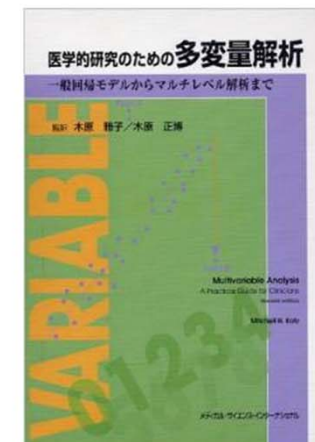
大橋靖雄. 2011
医歯薬出版
ASIN: B00I399SY4



矢野 栄二 (訳). 2013
篠原出版新社; 第2版
ISBN-13: 978-4884123727



木原雅子 (訳). 2013
メディカルサイエンスインターナショナル
ISBN-13: 978-4895927574



木原雅子 (訳). 2008
メディカルサイエンスインターナショナル
ISBN-13: 978-4895925532



- 因子分析
- 級内相関係数と分散分析の関係

因子分析

- 質問票が想定された通りの相関構造を示すかどうか？
- 例：HADS質問票の相関構造

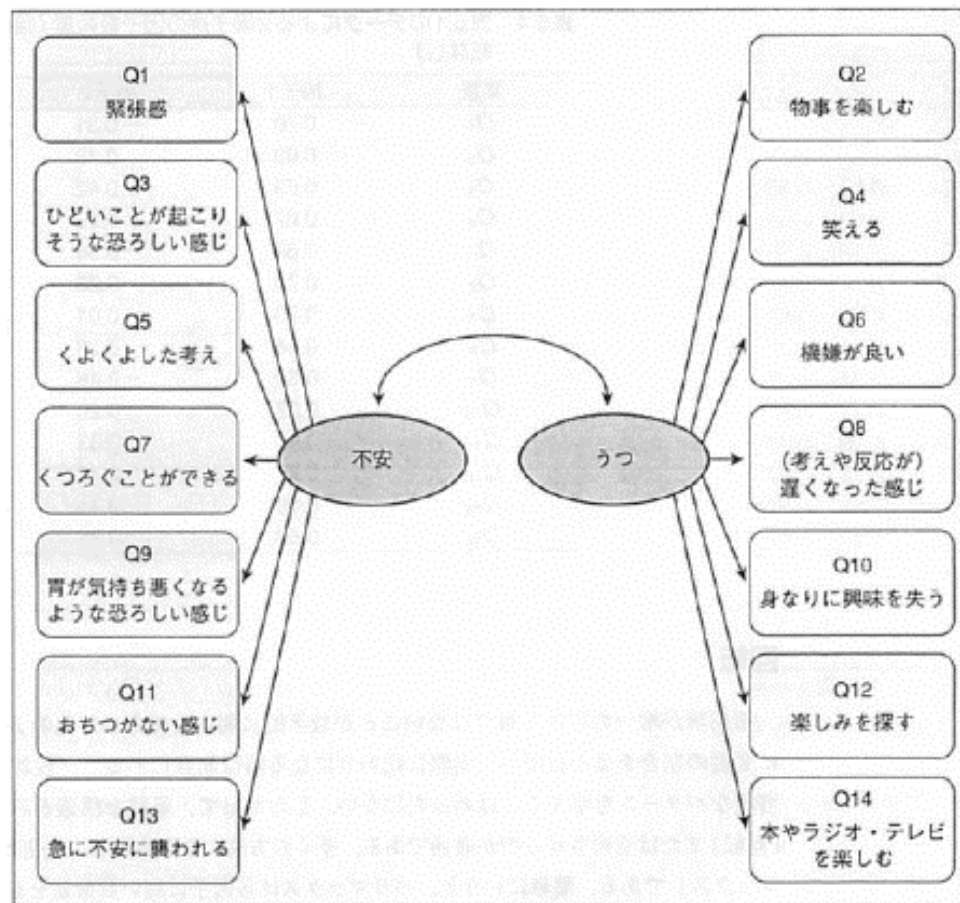


図 5.1 HADS 質問票の仮定された構造

- 奇数番号の項目が「不安」
- 偶数番号の項目が「うつ」

データから (隠れた) 潜在変数
を探すのが因子分析の目的

HADS質問票の相関行列

表 5.2 仮定された下位尺度不安 (奇数番項目) とうつ (偶数番項目) に一致するように表 5.1 を並べ替えた相関

	不安							うつ						
	Q ₁	Q ₃	Q ₅	Q ₇	Q ₉	Q ₁₁	Q ₁₃	Q ₂	Q ₄	Q ₆	Q ₈	Q ₁₀	Q ₁₂	Q ₁₄
Q ₁	1													
Q ₃	0.54	1												
Q ₅	0.56	0.60	1											
Q ₇	0.50	0.41	0.43	1										
Q ₉	0.49	0.58	0.52	0.38	1									
Q ₁₁	0.34	0.32	0.32	0.34	0.30	1								
Q ₁₃	0.54	0.60	0.56	0.43	0.57	0.38	1							
Q ₂	0.31	0.26	0.27	0.41	0.20	0.15	0.26	1						
Q ₄	0.34	0.36	0.36	0.43	0.28	0.18	0.33	0.47	1					
Q ₆	0.41	0.37	0.41	0.47	0.32	0.19	0.37	0.50	0.58	1				
Q ₈	0.28	0.23	0.27	0.33	0.17	0.16	0.23	0.52	0.33	0.37	1			
Q ₁₀	0.30	0.27	0.27	0.34	0.23	0.17	0.33	0.40	0.38	0.41	0.36	1		
Q ₁₂	0.33	0.32	0.31	0.44	0.26	0.17	0.30	0.59	0.54	0.52	0.44	0.45	1	
Q ₁₄	0.30	0.27	0.30	0.40	0.30	0.19	0.31	0.42	0.44	0.47	0.31	0.36	0.42	1

点線部分の方が実線部分よりも相関係数が低い傾向がある
(複雑な相関構造の場合はこの表を作っても構造がうまく捉えられない)

因子分析の流れ

1. 相関行列から項目毎に「固有値」を計算

- 固有値の合計 = 項目数 (HADS質問票の場合は14) = データの全変動
- 各項目の固有値が占める割合 = データ変動を説明する度合い
- 「固有値1以上」で因子数を決定するのが慣例的
 - HADS質問票の場合は2因子

表 5.3 表 5.1 の因子分析：固有値と説明される分散の割合

因子	固有値	差	説明される分散	
			割合	累積
1	5.84	4.08	0.42	0.42
2	1.76	0.94	0.13	0.54
3	0.82	0.07	0.06	0.60
4	0.75	0.07	0.05	0.65
5	0.68	0.07	0.05	0.70
6	0.61	0.04	0.05	0.75
7	0.57	0.06	0.04	0.79
8	0.51	0.07	0.04	0.83
9	0.44	0.00	0.03	0.86
10	0.43	0.03	0.03	0.89
11	0.40	0.00	0.03	0.92
12	0.40	0.03	0.03	0.95
13	0.37	0.00	0.03	0.98
14	0.37	—	0.02	1.00

2. 「因子負荷量」の計算

- 各因子における各項目の重要度（HADS質問票の場合は $14 \times 2 = 28$ 個）
- 因子負荷量の計算法
「各因子で、因子負荷量の値が高い変数の数を最小にする」
バリマックス回転法を使うのが一般的

表 5.5 表 5.4 から回転させた因子負荷行列：バリマックス回転

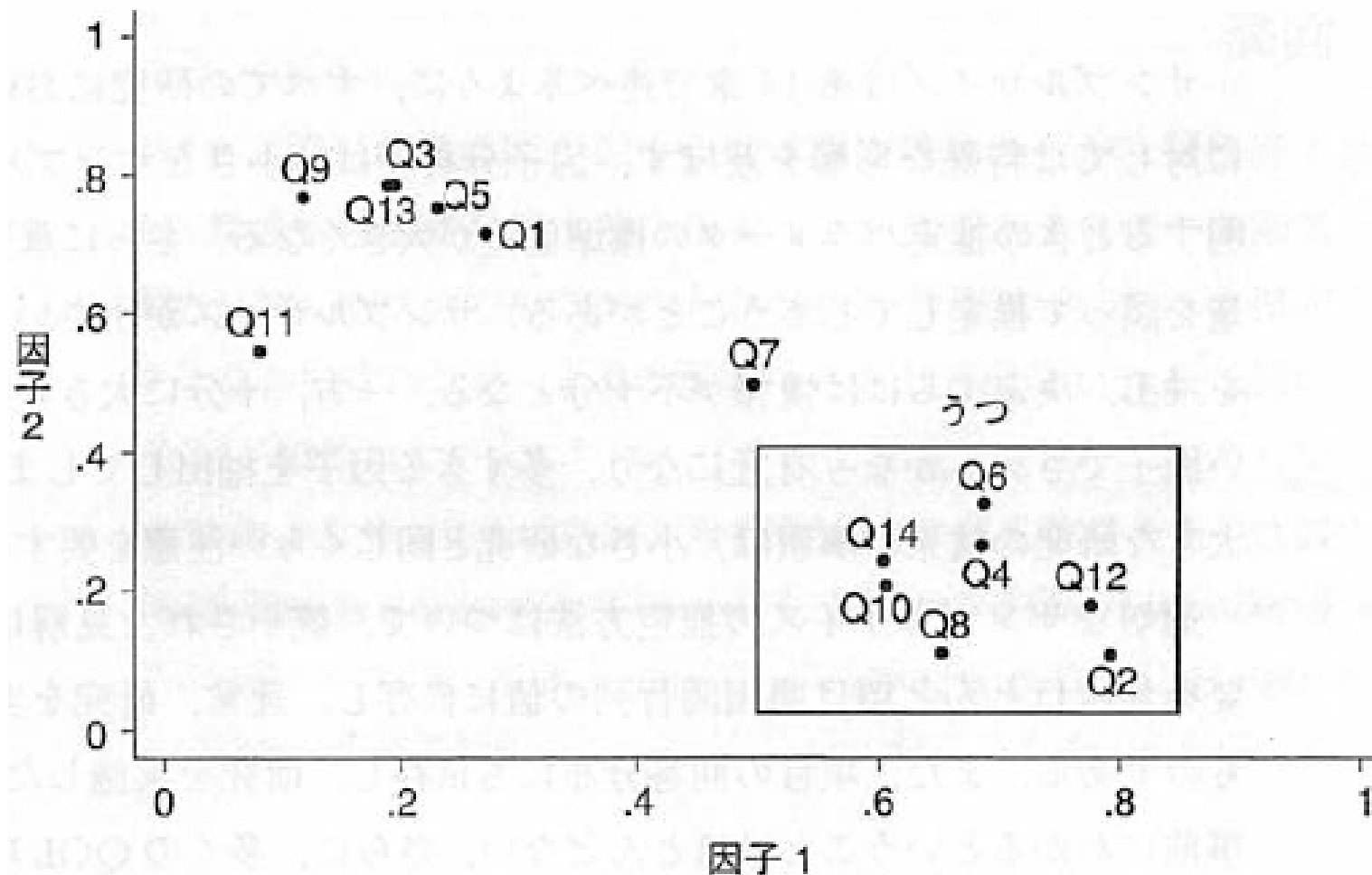
変数	因子1	因子2
Q_1	0.27	0.71
Q_2	0.79	0.11
Q_3	0.19	0.78
Q_4	0.68	0.26
Q_5	0.23	0.75
Q_6	0.68	0.32
Q_7	0.49	0.50
Q_8	0.65	0.11
Q_9	0.11	0.76
Q_{10}	0.60	0.20
Q_{11}	0.07	0.54
Q_{12}	0.77	0.18
Q_{13}	0.19	0.78
Q_{14}	0.60	0.24

因子負荷量の大きさに注目！

- 因子1が「うつ」（偶数項目）
- 因子2が「不安」（奇数項目）

3. 因子負荷量を視覚化

- 3因子以上ある場合はペアにして複数プロット作る
- HADS質問票では偶数項目が因子1に、奇数項目が因子2に分類された
 - Q7が「うつ」寄りの質問項目であることも分かる



級内相関係数 (ICC)の定義

■ もっともシンプルな(一元配置)分散分析モデル

患者*i*における1回目の測定値 = 集団平均 + 患者*i*の効果 + 誤差

患者*i*における2回目の測定値 = 集団平均 + 患者*i*の効果 + 誤差

⋮

患者*i*における*j*回目の測定値 = 集団平均 + 患者*i*の効果 + 誤差

何回目に測定したか？ は誤差に含まれる

誤差が対象によらず等しければ測定回数に関係なく測定値は等しい値になる

■ 分散でICC（信頼性係数； ρ ）を表現した式

$$\rho = \frac{\sigma_s^2}{\sigma_s^2 + \sigma_e^2} = 1 - \frac{\sigma_e^2}{\sigma_s^2 + \sigma_e^2}$$

- σ_s^2, σ_e^2 はそれぞれ患者効果(平均からのズレ)と誤差のバラツキ
- 誤差のバラツキがゼロに近づく とICCは1に近づく
- 誤差のバラツキが大きいとICCはゼロに近づく

ICCの一般的な定義

いずれも測定値の分散に占める
対象の効果の分散の割合を計算

■ 二元配置デザインを想定したモデルから導出

- 研究デザインによって計算方法が変わる
- 一元配置：対象内の繰り返し測定順によって影響がない場合

$$ICC = \frac{\hat{\sigma}_s^2}{\hat{\sigma}_s^2 + \hat{\sigma}_e^2}$$

- 二元配置：対象内の繰り返し測定順によって影響がある場合
 - 測定時期・評価者によって値が変動するような場合

$$x_{ij} = \mu + s_i + r_j + \text{誤差}$$

$$\sigma_x^2 = \sigma_s^2 + \sigma_R^2 + \sigma_e^2$$

測定時期・評価者の効果(r_j)とその分散(σ_R^2)

$$ICC = \frac{\text{対象の効果}(s_i)\text{の分散}(\hat{\sigma}_s^2)}{\text{対象の効果}(s_i)\text{の分散}(\hat{\sigma}_s^2) + \text{評価者の効果}(r_j)\text{の分散}(\hat{\sigma}_R^2) + \text{誤差の分散}(\hat{\sigma}_e^2)}$$

この式で推定したICCを「FleissのICC」と呼ぶ

営利目的でのご利用や無断転載はご遠慮ください

ICR臨床研究入門 <http://www.icrweb.jp>

補足：FisherのICCと分散分析の関係

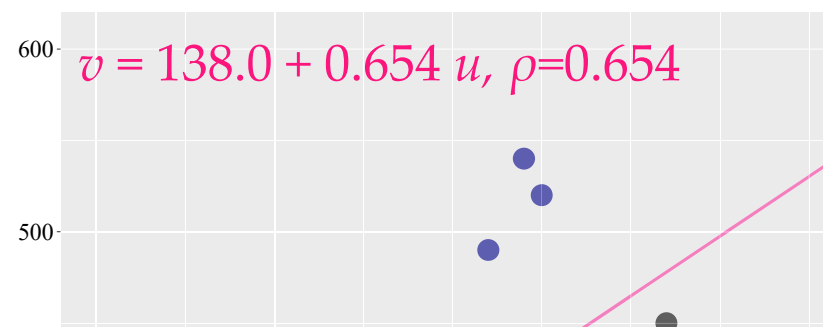
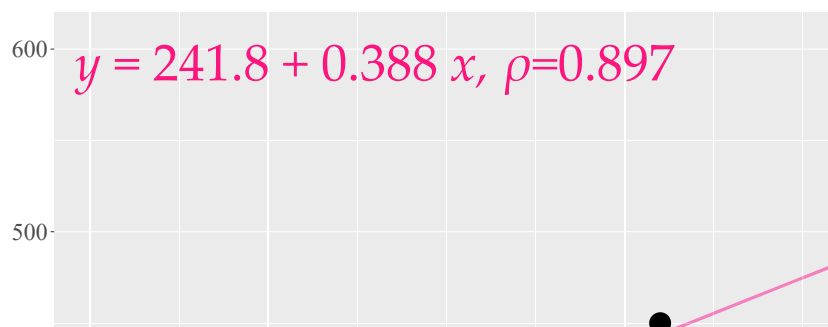
- FisherのICCと分散分析で計算したICCはサンプルサイズが十分に大きいとき一致
 - なぜか？
- FisherのICC計算と(一元配置)分散分析で用いる平方和の分解は(数学的に)リンクしている部分がある
 - 説明のため、 x と y を1回目/2回目の測定値、 (x, y) のペアを入れ替えた (y, x) を含む測定値を (u, v) とする

■ ピークフローデータ($n=7$)

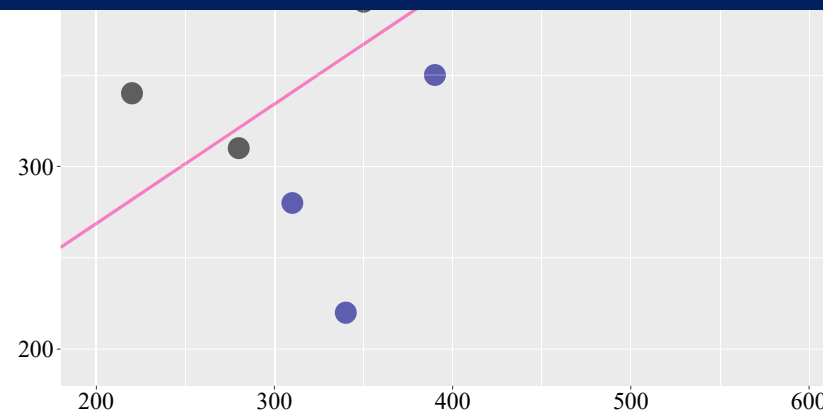
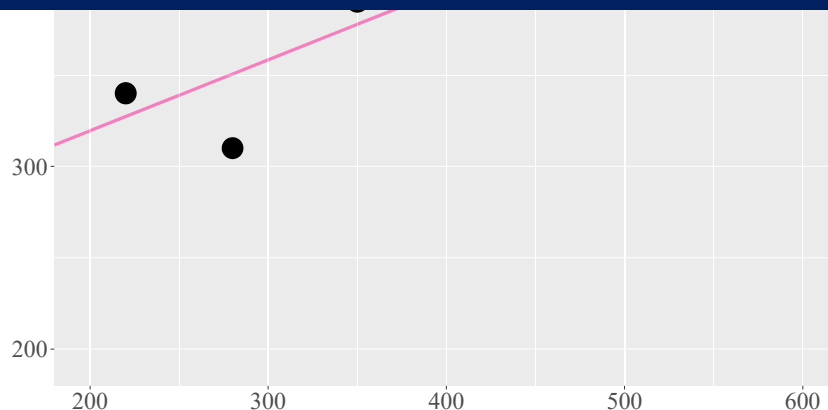
対象	1	2	3	4	5	6	7	平均
1回目(x)	280	540	400	490	520	220	350	400
2回目(y)	310	440	430	420	450	340	390	397.1
平均(m)	295	490	415	455	485	280	370	398.6

元データ : (x, y)

線対称データも含めたデータ : (u, v)



(x, y)のペアを入れ替えた(y, x)を含む(u, v)の相関係数 = FisherのICC



■ ピークフローデータ($n=7$)

対象	測定	データ
1	1回目	280
	2回目	310
2	1回目	540
	2回目	440
⋮		
7	1回目	350
	2回目	390
平均		398.6

(1) 左のデータに対して「対象の効果」を調べる
一元配置分散分析を行った結果

要因	平方和	自由度	平均平方
S (対象)	89771 (S_S)	6 ($= 7-1$)	14961.9
誤差	18800	7 ($= 7 \times (2-1)$)	2685.7
T (総和)	108571 (S_T)	13 ($= 7 \times 2-1$)	

(2) 平均平方から σ_s^2 , σ_e^2 が推定できる

$$\text{一元配置分散分析から計算したICC} = \frac{\hat{\sigma}_s^2}{\hat{\sigma}_s^2 + \hat{\sigma}_e^2} = 0.696$$

- 
- (u, v) それぞれの平方和 S_u と S_v は S_T に等しい(書き下せば当たり前)

$$S_u = S_v = S_T$$

- (u, v) の偏差積和 S_{uv} と対象間平方和 S_S には以下の関係

$$S_{uv} = 2S_S - S_T$$

- 以上からFisherのICCは

$$\text{FisherのICC} = \frac{S_{uv}}{\sqrt{S_u \times S_v}} = \frac{2S_S - S_T}{S_T}$$

■ 一元配置分散分析表より

$$2\hat{\sigma}_S^2 + \hat{\sigma}_e^2 = \frac{S_S}{n-1}, \quad \hat{\sigma}_e^2 = \frac{S_e}{n}$$

■ 以上の関係式から、

$$\text{一元配置分散分析から計算したICC} = \frac{nS_S - (n-1)S_e}{nS_S + (n-1)S_e}$$

■ サンプルサイズ(n)が十分に大きいとき、 n と $n-1$ がほぼ同じ値と近似できるので

$$\text{一元配置分散分析から計算したICC} = \frac{nS_S - (n-1)S_e}{nS_S + (n-1)S_e} \approx \frac{2S_S - S_T}{S_T}$$

これは前ページのFisherのICCと同じ