

# 保健統計学 多変量解析の基礎 重回帰分析

米倉佑貴(yyonekura@slcn.ac.jp)

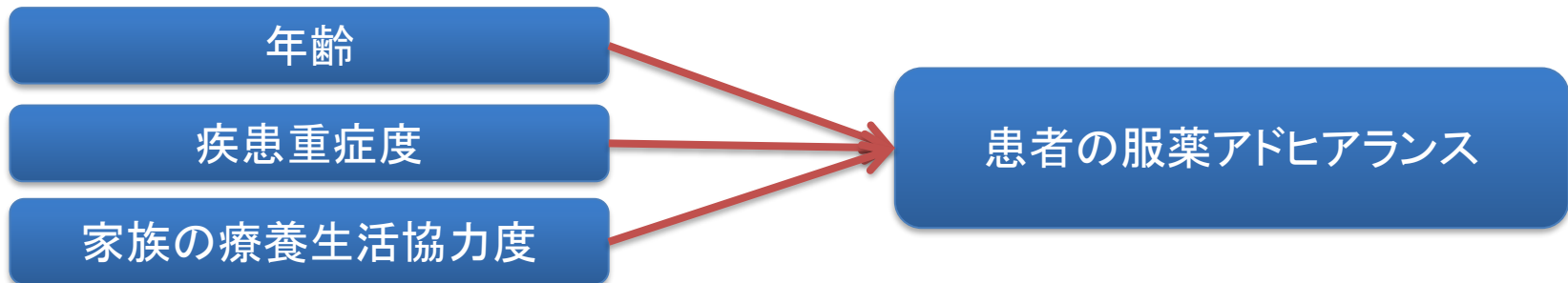
聖路加国際大学大学院

# ポイントとキーワード

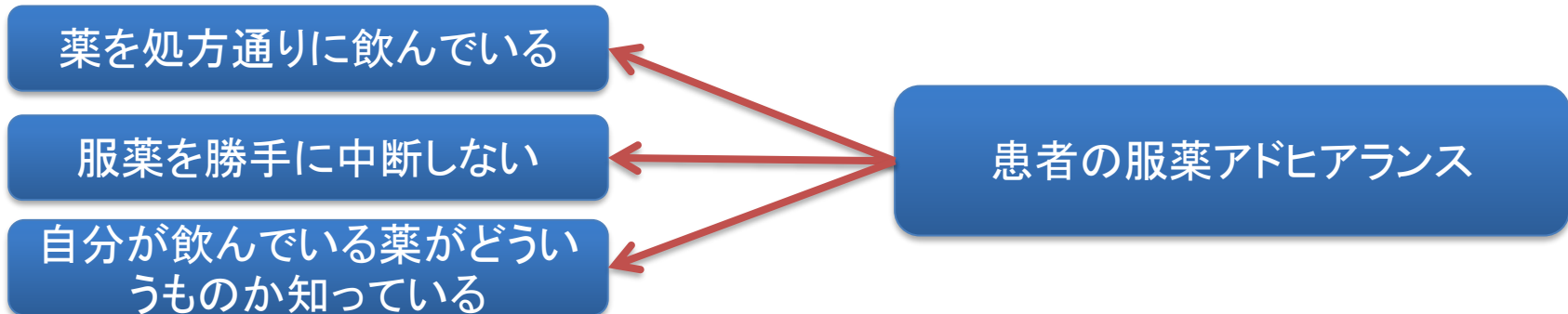
- 変数間の関係
  - 交絡, 媒介, 修飾, 抑制, 歪曲
- 多変量解析の基礎
  - 重回帰分析, 一般線形モデル
    - 最小二乗法
    - 主効果と交互作用効果
    - ダミー変数
    - 偏回帰係数, 標準化偏回帰係数
    - 決定係数
    - 多重共線性
    - 回帰診断

# 多変量解析とは？

- 3つ以上の変数の関連性を明らかにする方法
  - 説明型と分類型の2つに分けられる
- **説明型**: 多くの変数で一つの行動や状態などを説明する方法  
→ 重回帰分析、一般線形モデル、ロジスティック回帰分析、一般化線形モデルなど



- **分類型**: 多くの変数をまとめたり、分類したりする方法  
→ 主成分分析、因子分析など



# いろいろな多変量解析

## 説明型

一般線形モデル(General Linear Model)→目的変数が正規分布  
分散分析(ANOVA)  
共分散分析(ANCOVA)  
(重) 回帰分析  
(Multiple) Regression analysis

一般化線形モデル(Generalized Linear Model)→目的変数が指数分布族  
ロジスティック回帰分析  
(Logistic regression)  
プロビット回帰分析  
(Probit regression)  
ポアソン回帰(Poisson regression)

## 分類型

主成分分析(Principal Component Analysis)→合成変数の作成

因子分析(Factor Analysis)→多くの変数に共通する因子をみつける

クラスター分析(Cluster Analysis)→得られたデータからケース(対象者)を分類

判別分析(Discriminant analysis)→事前に持っているデータをもとに、得られたケースを分類

# 多変量解析はなぜ必要なのか？

- 説明型

- 注目する要因以外の条件を揃えるために使う
  - 多変量解析で交絡を除去する
- 多くの要因で説明したほうが、複雑な行動や特性や状態をうまく説明できる
  - 2変数の分析では表現できない複雑な関連性を明らかにすることができる
  - ただし、あまり要因を多くし過ぎると解釈が難しくなる

- 分類型

- 多くの変数の情報を少ない情報に集約してわかりやすくする
- 質問紙や検査などでは直接測定できないものを統計モデルにより推定する
  - 知能, 幸福, 健康, 生活の質, 愛, 性格など

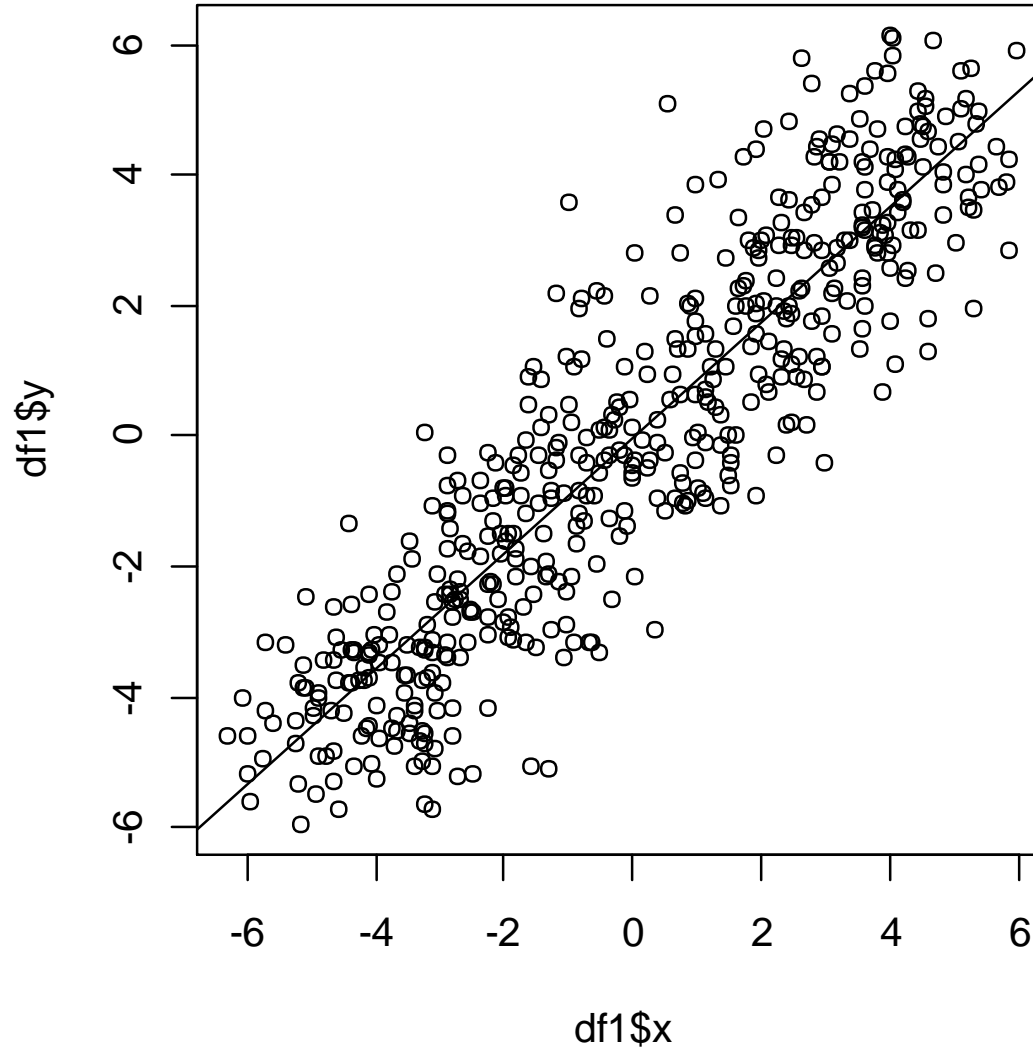
# いろいろな変数間の関係

X: 説明変数, Y: 目的変数, T: 第3変数

第3変数(T)と説明変数(X)の関係	2変量間の関係	第3変数で制御した時のXとYの関連性	呼び方
	XとYが関連 TとX, TとYが関連	消失	交絡
	XとYが関連	消失もしくは弱まる	媒介
	XとYが関連	Tの値によって関連性が異なる	修飾
	XとYは無関係	関連性が出現	抑制
	XとYが関連	関連性の方向が逆になる	歪曲

# 交絡・媒介の視覚的イメージ

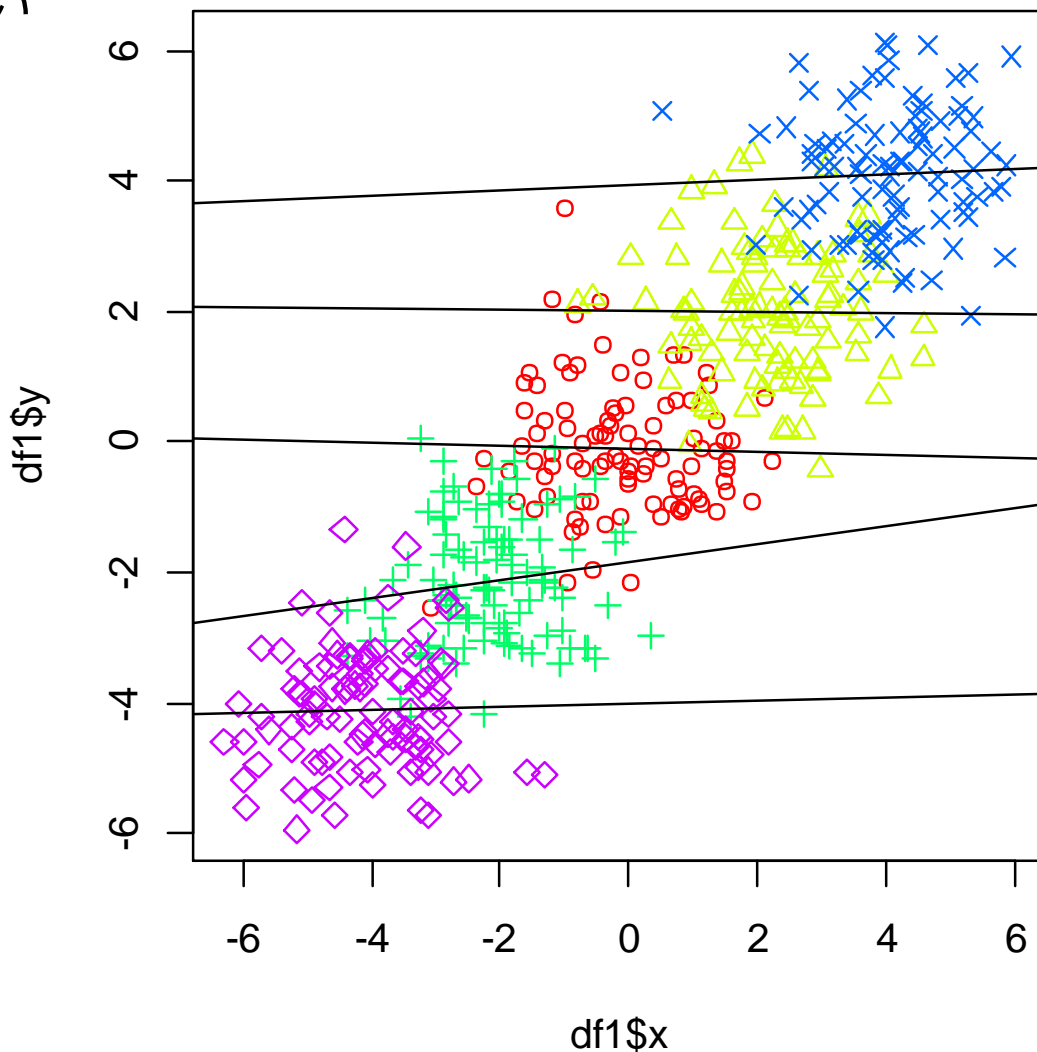
二変量では関連が認められる  
→xが大きいとyも大きい



# 交絡の視覚的イメージ

交絡要因または媒介要因(この場合は色分けしているグループ)で調整すると  
関連性が消失, または弱まる

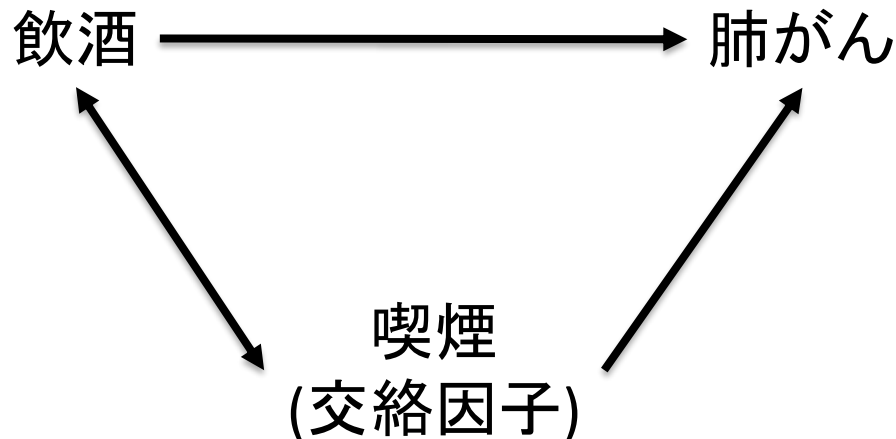
→グループごとにみるとxとyには関連性がない(相関係数が0に近い, 回帰  
係数が0に近い)





# 交絡

- 注目している結果と原因の関係がそれ以外の第3の要因によって攪乱されること
- 交絡因子
  - 原因(曝露要因)と結果(疾患の発生など)の両方に独立に関連する要因

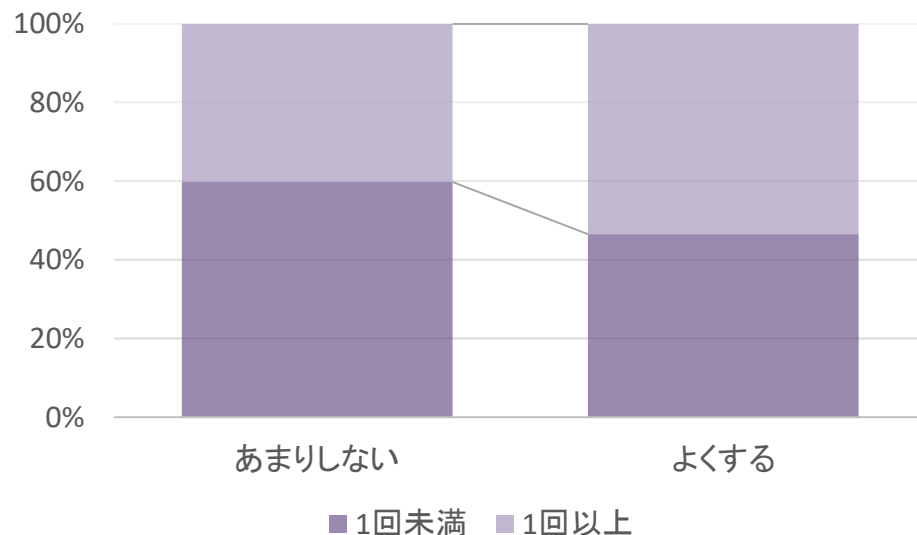


# 囲碁・将棋は健康に悪い？ 架空のデータから

## 月の通院回数

	1回未満	1回以上
囲碁・将棋をあまりしない	664(59.8%)	447(40.2%)
囲碁・将棋をよくする	434(46.6%)	498(53.4%)

囲碁・将棋をよくする人の方が  
通院回数が多い  
→囲碁・将棋は健康に悪いと  
言えるだろうか



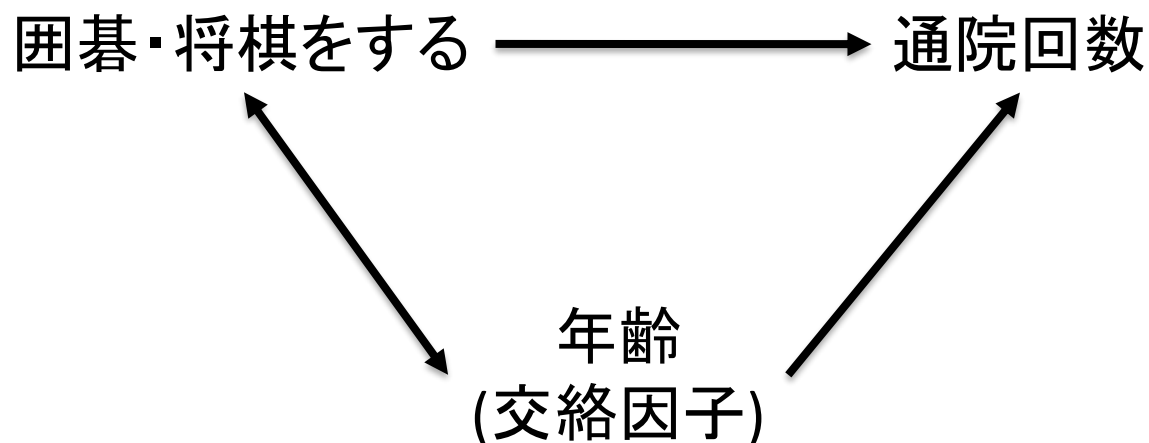
# 年齢別に見ると

		月の通院回数			
		1回未満		1回以上	
20-39歳	囲碁・将棋をあまりしない	312	(74.1%)	109	(25.9%)
	囲碁・将棋をよくする	55	(66.3%)	28	(33.7%)
40-64歳	囲碁・将棋をあまりしない	302	(60.4%)	198	(39.6%)
	囲碁・将棋をよくする	282	(57.3%)	210	(42.7%)
65-74歳	囲碁・将棋をあまりしない	35	(29.7%)	83	(70.3%)
	囲碁・将棋をよくする	70	(31.7%)	151	(68.3%)
75歳以上	囲碁・将棋をあまりしない	15	(20.8%)	57	(79.2%)
	囲碁・将棋をよくする	27	(19.9%)	109	(80.1%)

囲碁・将棋をよくする人とあまりしない人で月1回以上通院する人の割合の違いが小さくなった

→年代が同じであれば囲碁・将棋をするかしないかでは通院回数は変わらない  
 →通院回数の違いを説明する要因は囲碁・将棋が趣味かどうかではなく、年齢の違い

# 囲碁・将棋と通院回数の例に置き換えると



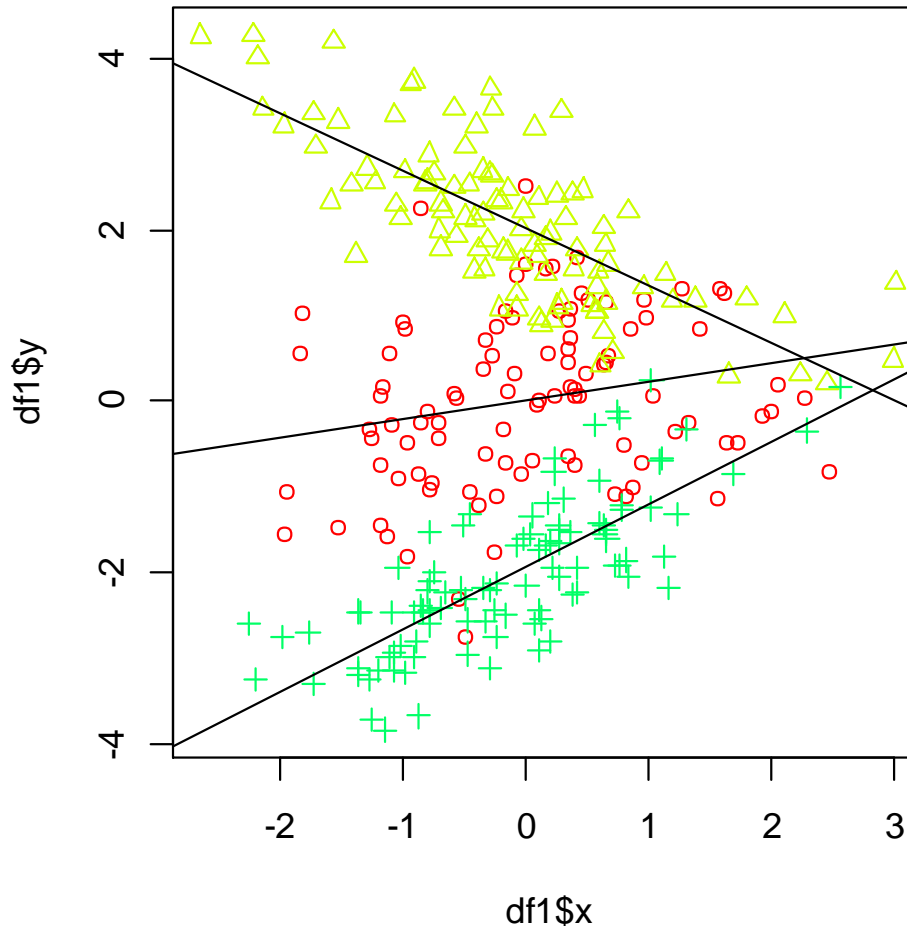
年齢が交絡因子になる

# 交絡因子への対策

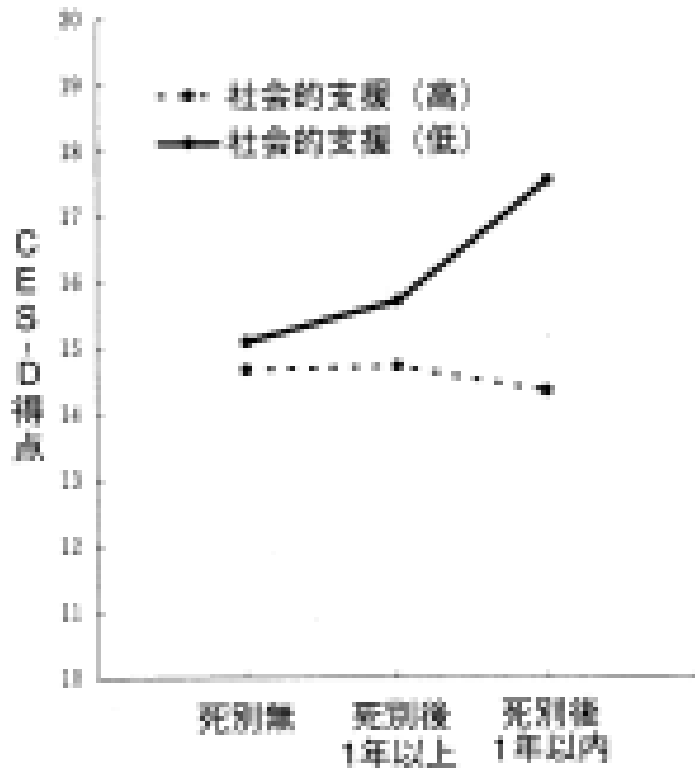
- 研究デザインでの対処
  - 研究対象者の制限
  - マッチング
  - 無作為化
- 解析段階での対処
  - マッチング
  - 層化
  - **多変量解析**

# 修飾の視覚的イメージ

修飾する要因の違いによって目的変数と説明変数の関連性が異なる  
→グループ(形と色で分けてある3グループ)によってxとyの関連性が異なる.  
△グループではxとyは負の相関, ○グループでは弱い正の相関(もしくはは無関係),  
+グループでは正の相関がみられる



# 修飾の例

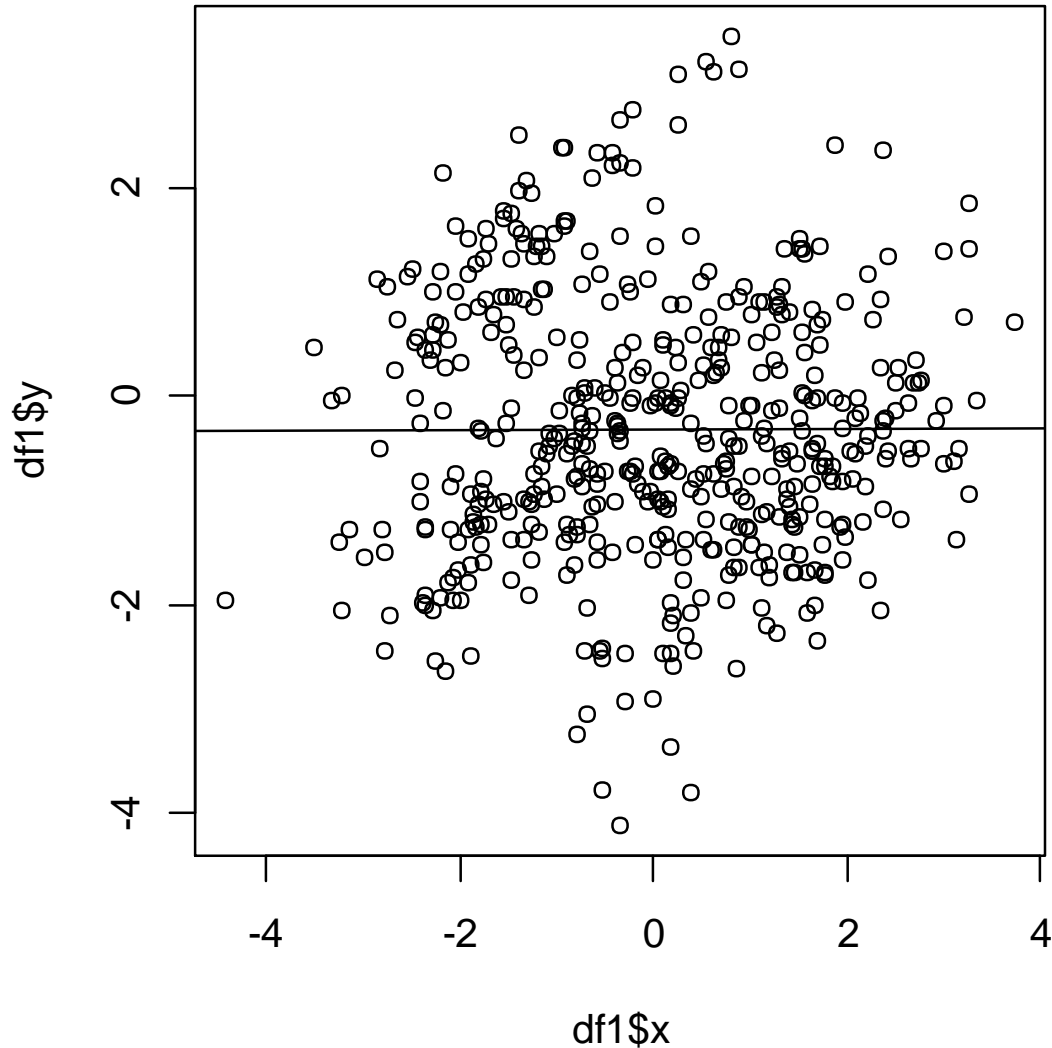


- 配偶者と死別した高齢者の抑うつ度と死別後の期間，社会的支援の関連
- 社会的支援が高い群では，死別後の期間が短くても抑うつ度が低く抑えられているが，社会的支援が低い群では，死別なしの対象者に比べて，死別後の期間が短い群で抑うつ度が高い

# 抑制の視覚的イメージ

二変量では関連が認められない

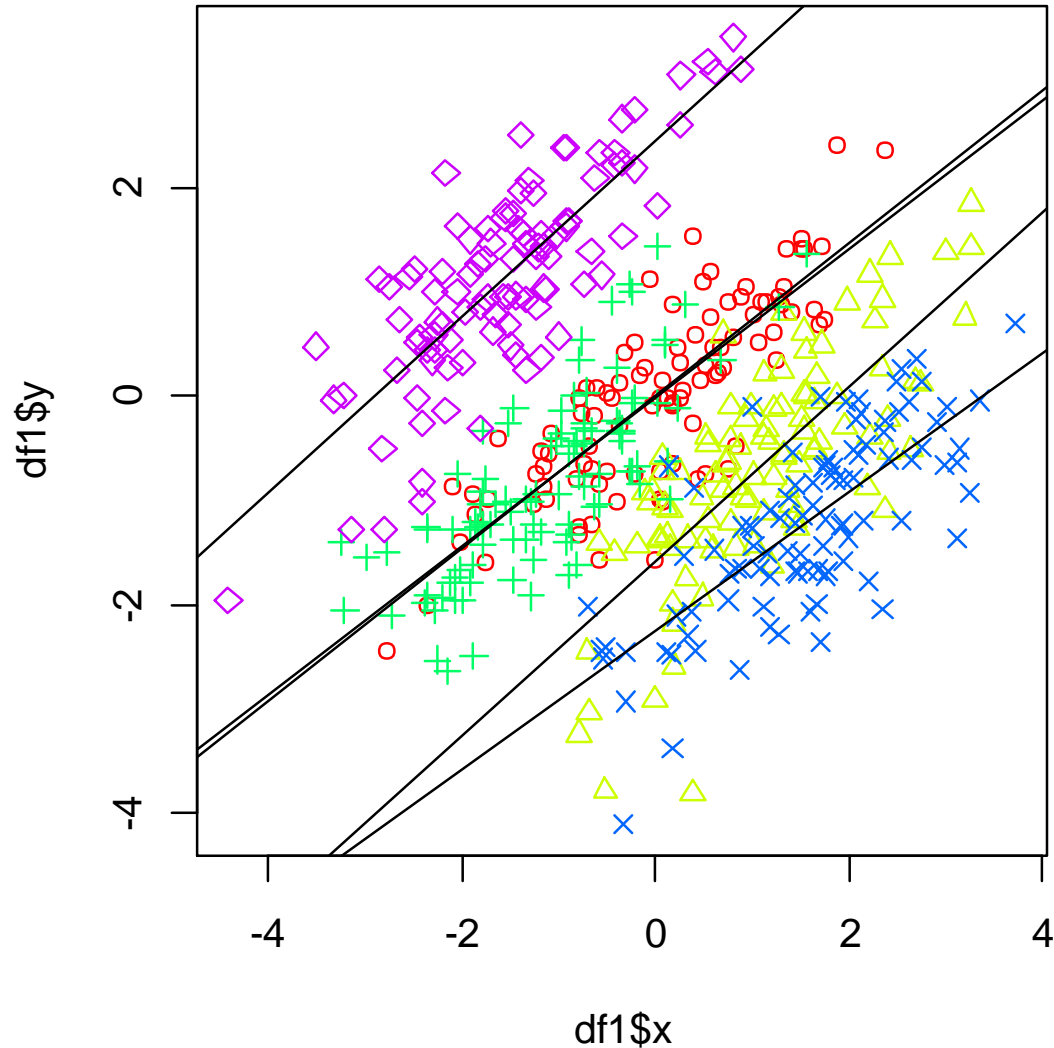
→回帰直線の傾きがほぼゼロ(x軸と平行)





# 抑制の視覚的イメージ

抑制変数(この場合は色分けしているグループ)で分けて分析すると関連が出現する



# 抑制の例

**TABLE 1.** CHARACTERISTICS OF INJURIES SUSTAINED BY CASE PATIENTS AND CONTROLS.

RISK FACTOR	CASE PATIENTS		CONTROLS		CRUDE ODDS RATIO (95% CI)*	P VALUE†
	NO. OF PATIENTS‡	% WITH RISK FACTOR	NO. OF PATIENTS‡	% WITH RISK FACTOR		
Large-gauge (<18) hollow-bore needle	27	15	488	1.2	14 (4.9–39)	0.001
Deep injury	33	52	675	6.8	15 (8.0–26)	<0.001
Visible blood on device	32	84	632	35	10 (4.6–23)	<0.001
Procedure involving needle in artery or vein	33	73	669	31	5.9 (2.9–12)	<0.001
Emergency procedure	33	12	661	2.4	5.6 (2.0–16)	0.012
Use of gloves	32	78	679	78	1.0 (0.4–2.4)	1.0
AIDS in source patient	33	82	676	70	1.9 (0.8–4.6)	0.18
Terminal illness in source patient§	27	48	349	16	4.8 (2.3–10)	<0.001
Postexposure use of zido- vudine	33	27	679	36	0.7 (0.3–1.4)	0.35

- 針刺し事故によるHIV感染についてのケースコントロール研究
- 二変量の分析ではジドブジン(逆転写酵素阻害薬)の曝露後予防内服の曝露オッズ比は0.7で有意ではない

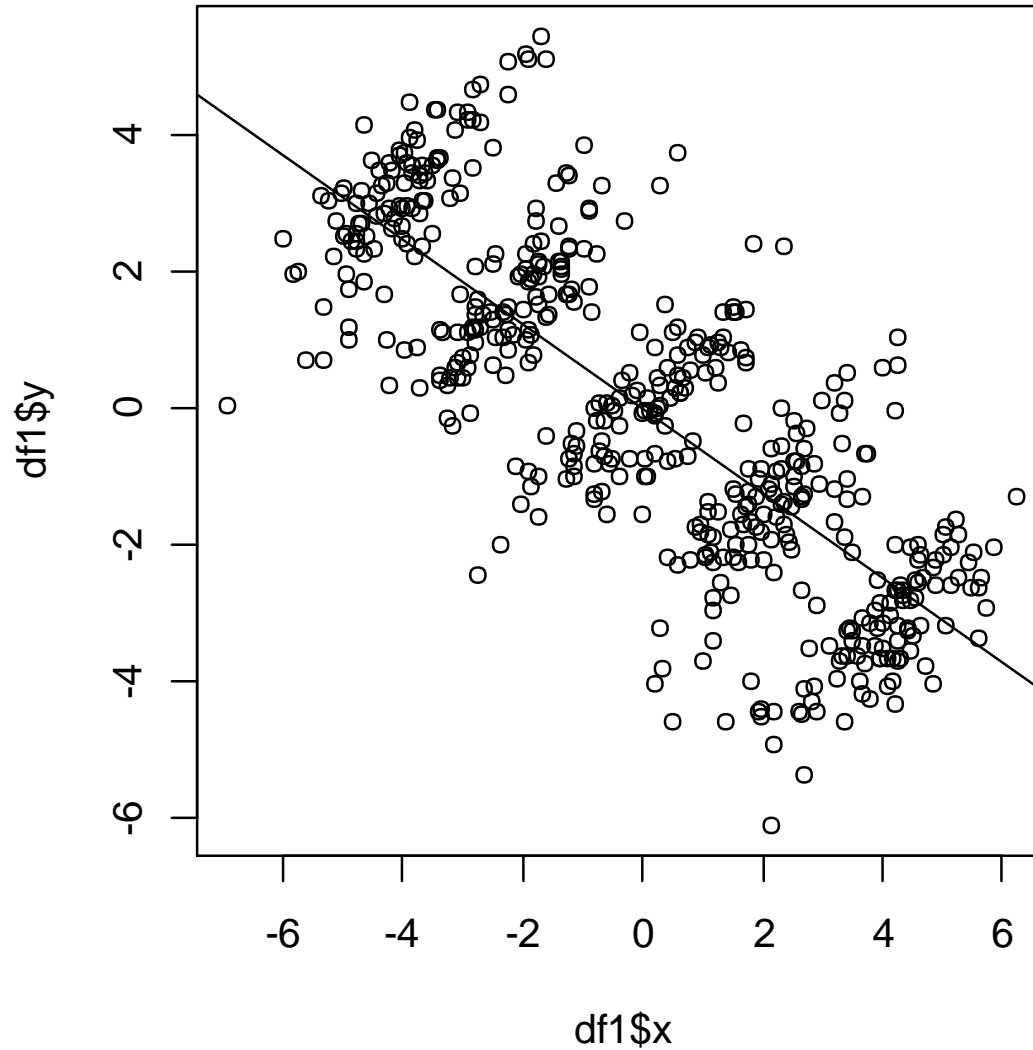
**TABLE 2. LOGISTIC-REGRESSION ANALYSIS OF RISK FACTORS FOR HIV TRANSMISSION AFTER PERCUTANEOUS EXPOSURE TO HIV-INFECTED BLOOD.**

RISK FACTOR	U.S. CASES*	ALL CASES†
	adjusted odds ratio (95% CI)‡	
Deep injury	13 (4.4–42)	15 (6.0–41)
Visible blood on device	4.5 (1.4–16)	6.2 (2.2–21)
Procedure involving needle in artery or vein	3.6 (1.3–11)	4.3 (1.7–12)
Terminal illness in source patient§	8.5 (2.8–28)	5.6 (2.0–16)
Postexposure use of zidovudine	0.14 (0.03–0.47)	0.19 (0.06–0.52)

- 多変量調整をすると曝露後予防内服のオッズ比は0.19で有意になった
- 深刻な事故であるほど曝露後予防内服をする可能性が高かったため、深刻さを制御しない二変量の分析では予防内服は予防に効果がないようにみえていた
- 事故の深刻さを制御することで、曝露後予防内服が感染の予防に効果的である可能性が示された

# 歪曲の視覚的イメージ

二変量では負の関連



# 歪曲の視覚的イメージ

多変量では正の関連といったように、二変量で見られる関連性と多変量で見られる**関連性の符号が異なる**

二変量が正の関連で多変量では負の関連ということもある

