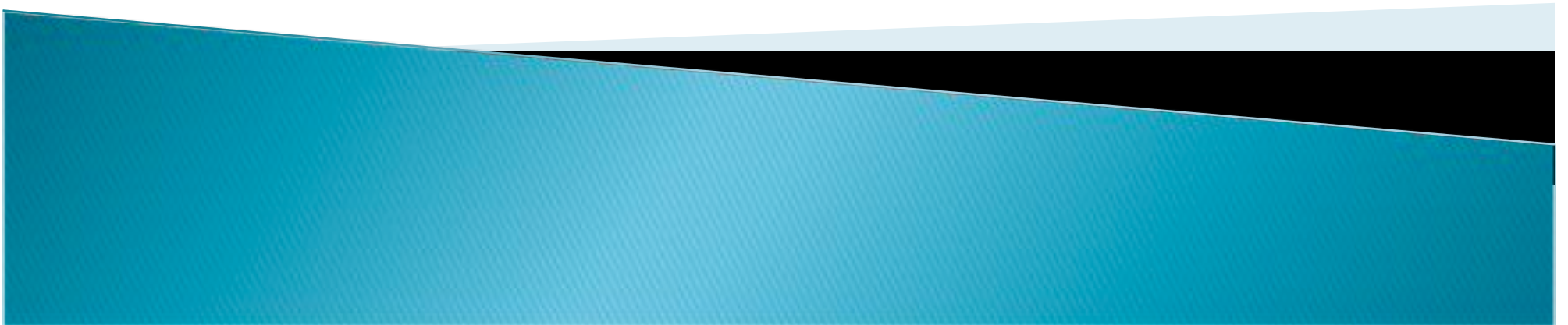



# 多变量解析 (Multivariate Analysis)

2010/11/19

潮雅之



# 今日の内容

- ▶ どのときに使用するか?
  - ▶ PCA, CA
  - ▶ DCA
  - ▶ RDA, CCA
  - ▶ その他の多変量解析
  
  - ▶ PCAの実践
  - ▶ RDAの実践
  - ▶ Permutation test, Partial RDA, Variation partitioningについて
- 

# どういうときに使用するか?

## ▶ 生物群集の組成データ

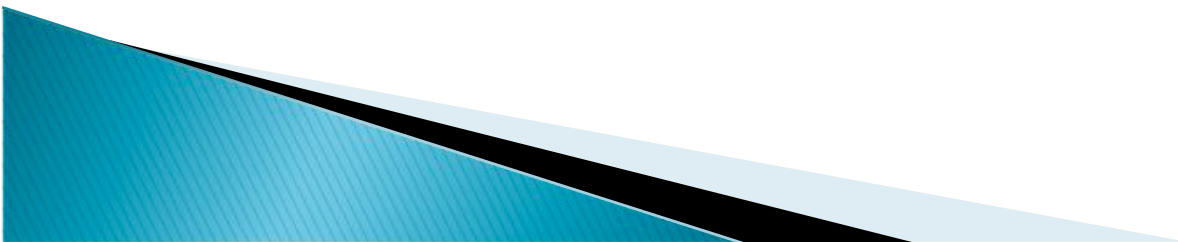
生物種	調査地 1	調査地 2	調査地 3	調査地 4
種 A	100	200	100	80
種 B	15	10	10	10
種 C	10	20	10	15
種 D	50	20	30	30
...	...	...	...	...

- サイト間で生物群集の組成に違いがあるか
- どのサイトが最も'特殊な'群集組成か
- 群集組成の環境データの関係性を知りたい

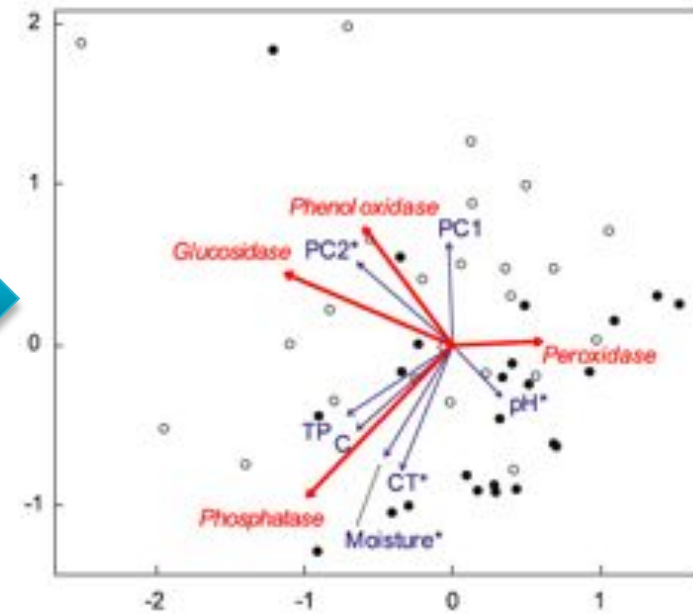


# どういうときに使用するか？

- ▶ 多変量のデータを要約して、1枚の図としてみせることができる。
- ▶ あまりはっきりしたことは言えないが（例えば、メカニズムがどうか）、データパターンを把握するのに役立つ。
- ▶ 特に多変量を扱うことが多い社会科学や生態学分野でよく使われているようである。




The image shows a screenshot of an Excel spreadsheet with a large table of data. The spreadsheet has a standard Excel interface with a menu bar at the top (File, Edit, View, Insert, Format, Tools, Data, Window, Help) and a toolbar. The data table consists of many columns and rows, with some cells containing numerical values and others containing text. The spreadsheet is titled 'Tpd6.csv'.

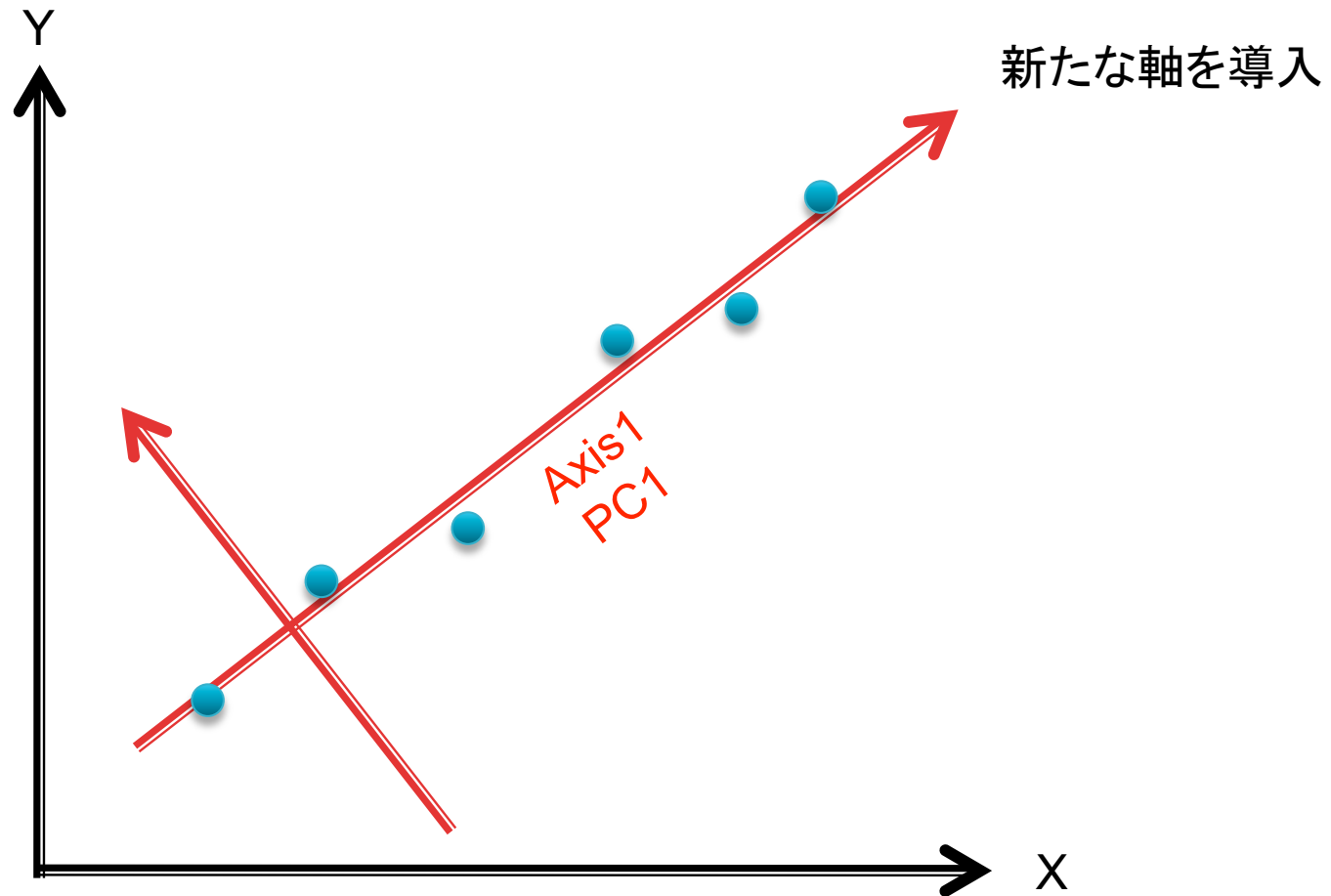


# 多変量解析のいろいろ

# PCA/CA

- ▶ 多変量のデータセットの要約(座標付け [ordination]、次元を減らす)をしたい。
  - ▶ 主成分分析  
Principle Component Analysis (PCA)
  - ▶ 対応分析  
Correspondence Analysis (CA)
  - ▶ 生物群集の構造解析のために発展してきた。
- 

# 次元を減らす、という考え方



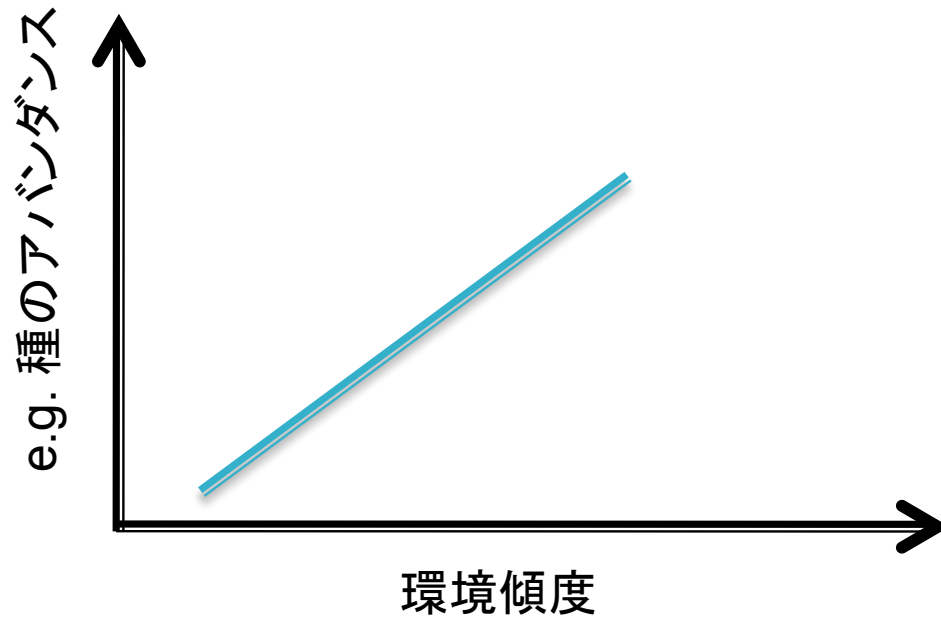
2変量データが1変量データに！

... ただし、元のデータを全て表している訳ではない



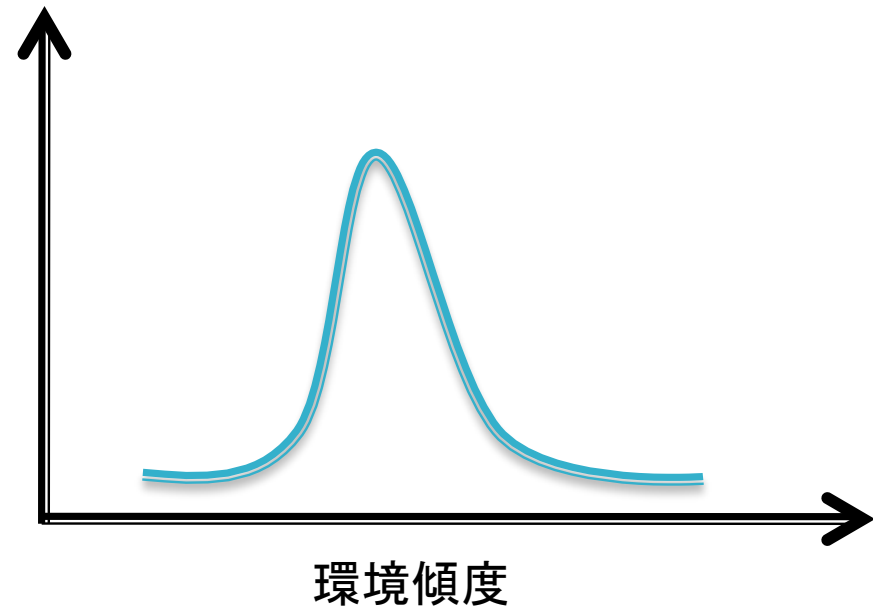
# PCA/CA

▶ PCA



直線的な関係

▶ CA



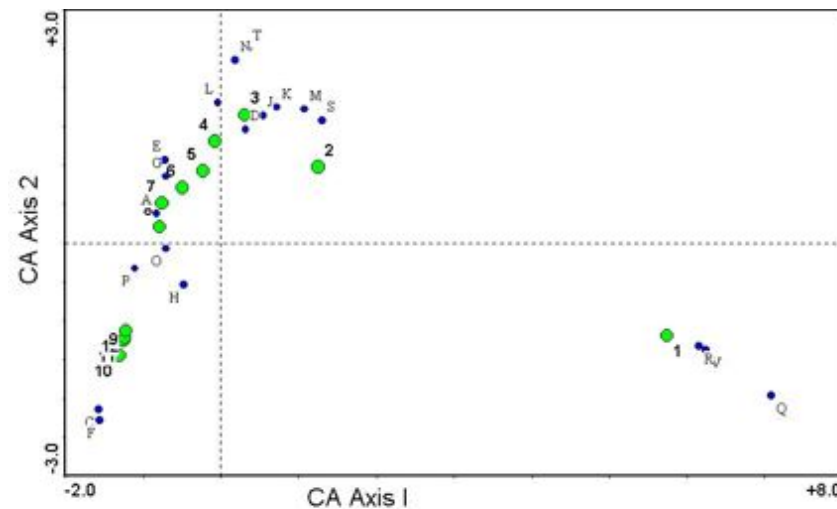
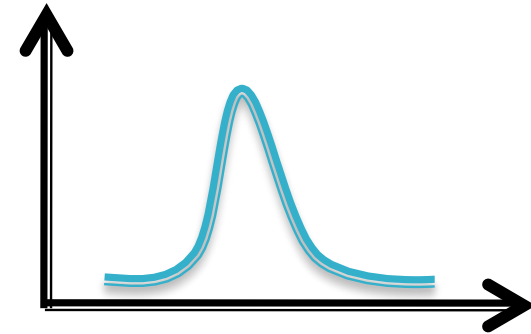
一山型の関係



# DCA

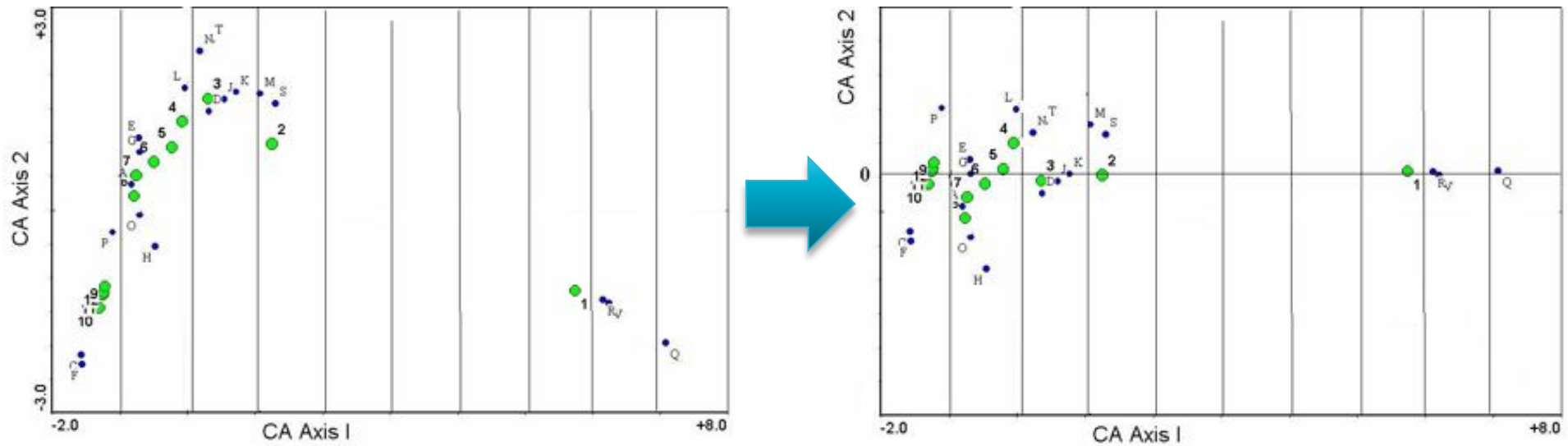
- ▶ 除歪対応分析

Detrended Correspondence Analysis (DCA)



馬蹄効果 (horseshoe effect)

# DCA

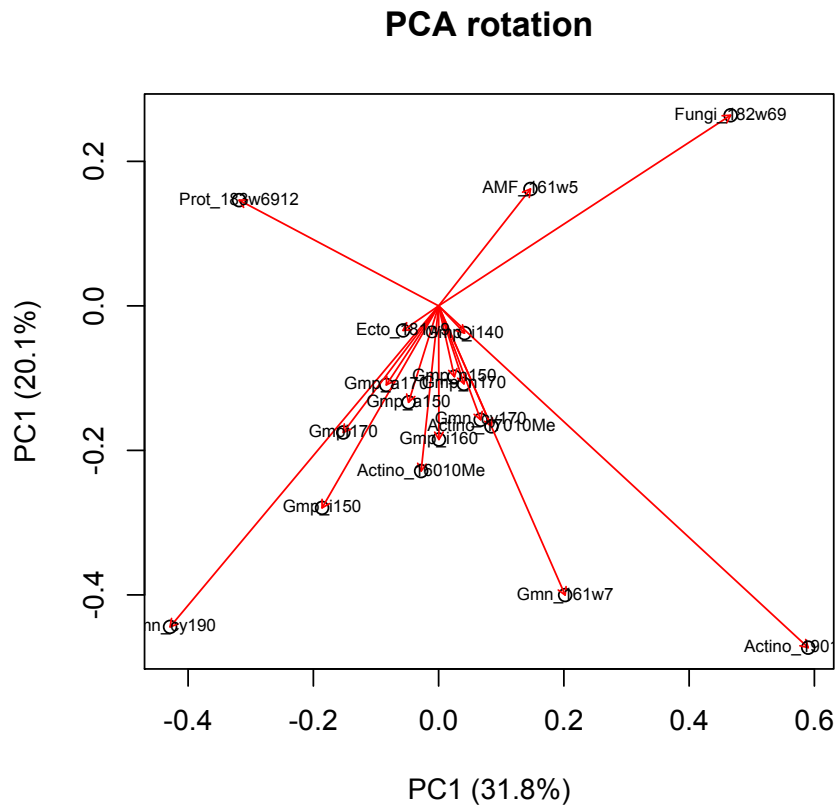


この操作で馬蹄効果を取り除くことができる

# PCA or CA/DCA ?

- ▶ DCAを一度やってみて、gradient lengthという値をチェックする。
- ▶ gradient length > 4  
環境傾度が広い→ 一山型を仮定した解析
- ▶ gradient length < 4  
環境傾度が狭い→ 直線を仮定した解析

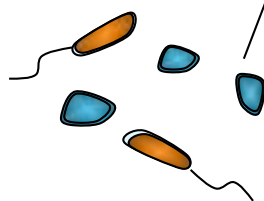
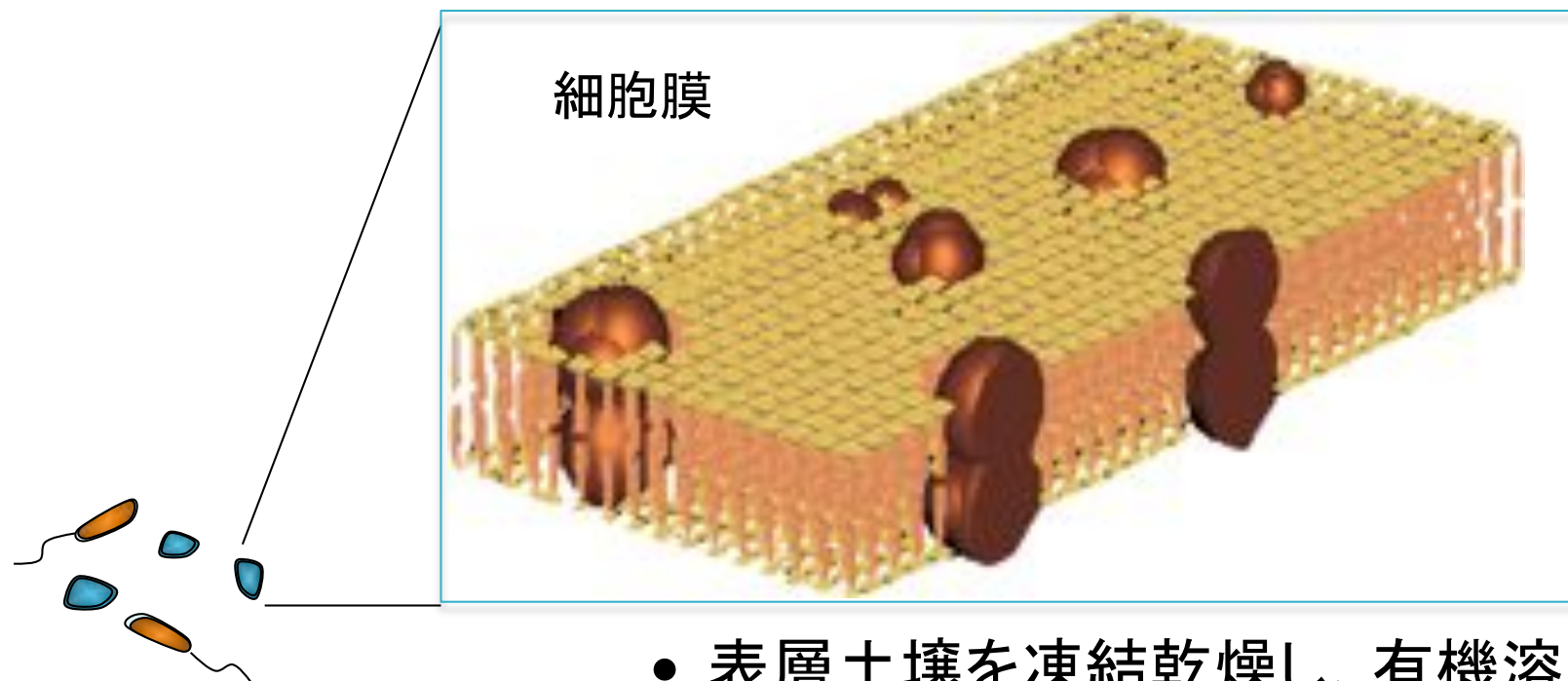
# RでPCAを実践



- ▶ 土壌微生物群集の組成データを解析。
- ▶ 土壌微生物群集の組成は土壌から抽出された脂質濃度を指標として解析

# 脂質データ

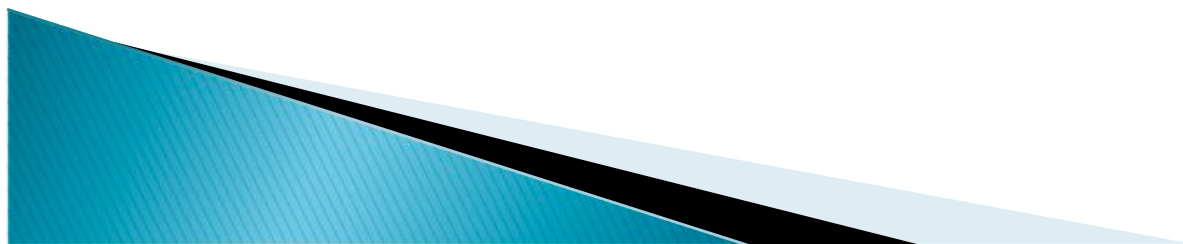
Lipid biomarker



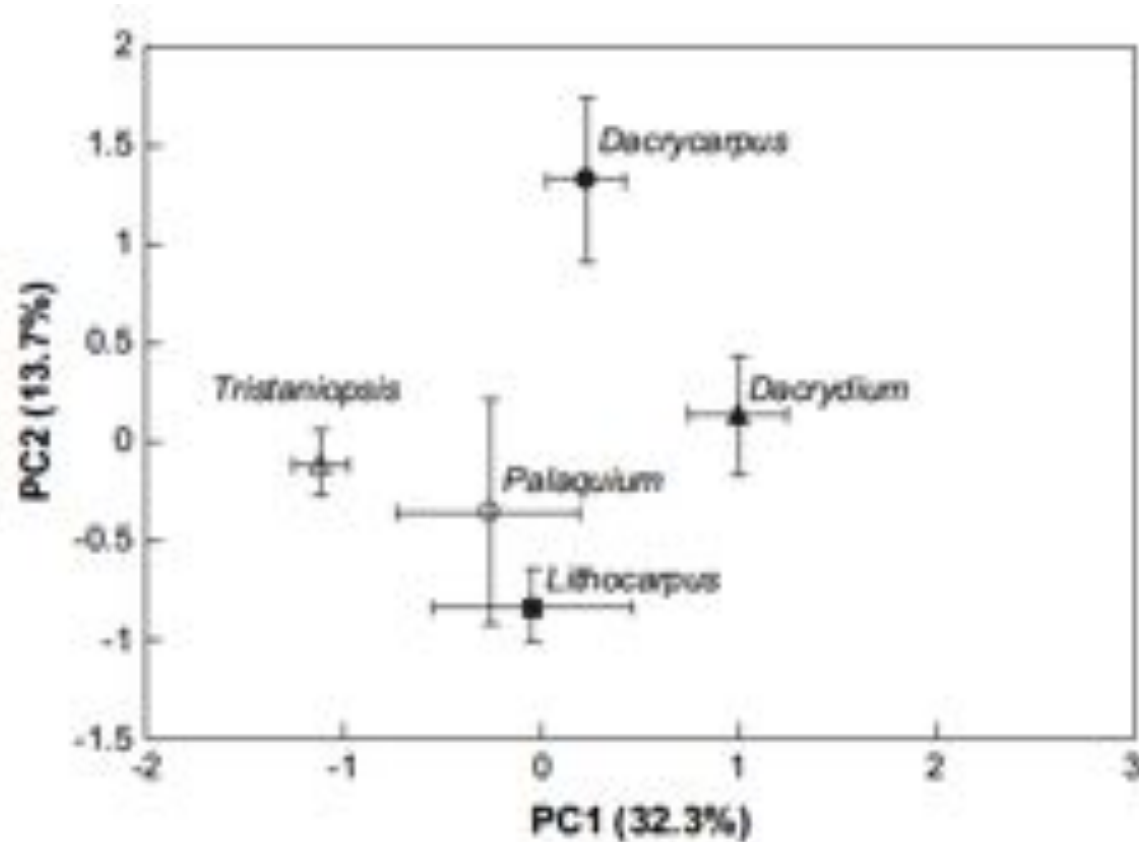
微生物細胞

- 表層土壌を凍結乾燥し、有機溶媒で抽出
- GCで分析
- 検出されたピークのうち、分類群特異的な脂肪酸を同定

脂質濃度 (nmol g <sup>-1</sup> )				
脂質名称	由来	サンプル 1	サンプル 2	サンプル 3
18:2 $\omega$ 6,9	真菌類	100.0	200.0	100.0
15:0	G+細菌	15.0	10.0	10.0
16:1 $\omega$ 7	G-細菌	50.0	20.0	30.0
17:0 10Me	放線菌	10.0	20.0	10.0
16:1 $\omega$ 5	AM菌根菌	50.0	20.0	30.0
18:1 $\omega$ 9	外生菌根菌	10.0	20.0	10.0
18:1 $\omega$ 6,9,12	原生動物	50.0	20.0	30.0
...	...	...	...	...



# RでPCAを実践



- ▶ 得られた主成分を各グループ間で比較して、全体で群集組成が違ふ、などと議論する。

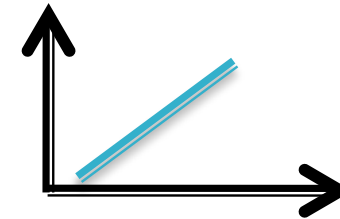


# RDA/CCA

- ▶ 環境条件と群集構造の関係を知りたい
- ▶ 群集構造 ~ 環境条件(説明要因)

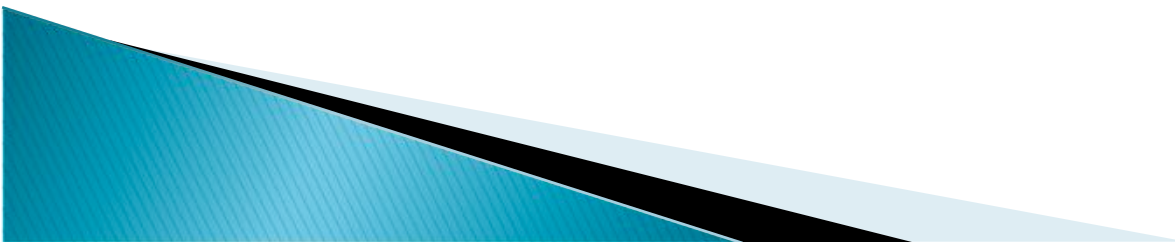
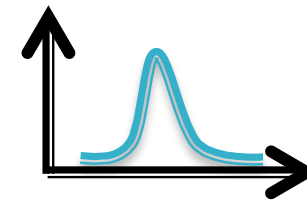
- ▶ 冗長性解析

Redundancy Analysis (RDA)



- ▶ 正準対応分析

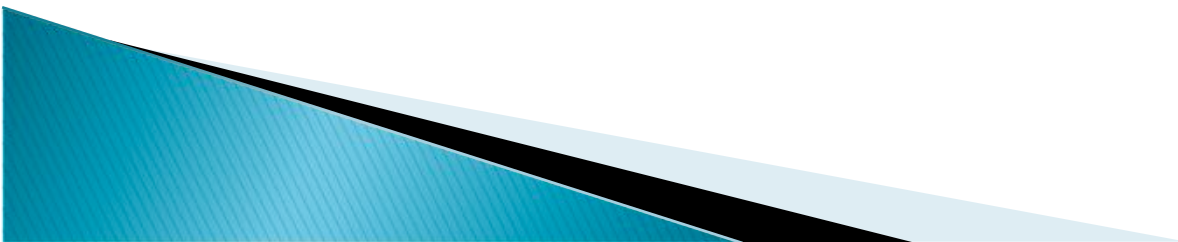
Canonical Correspondence Analysis (CCA)



# RDA/CCA



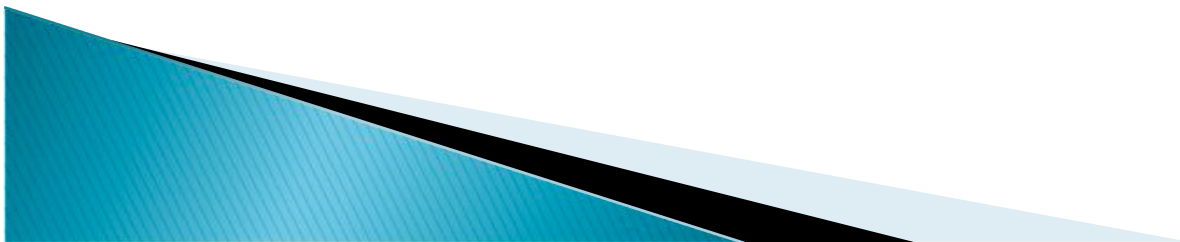
どの変数が有為な効果を持っているか知りたい。  
その場合は→Permutation test



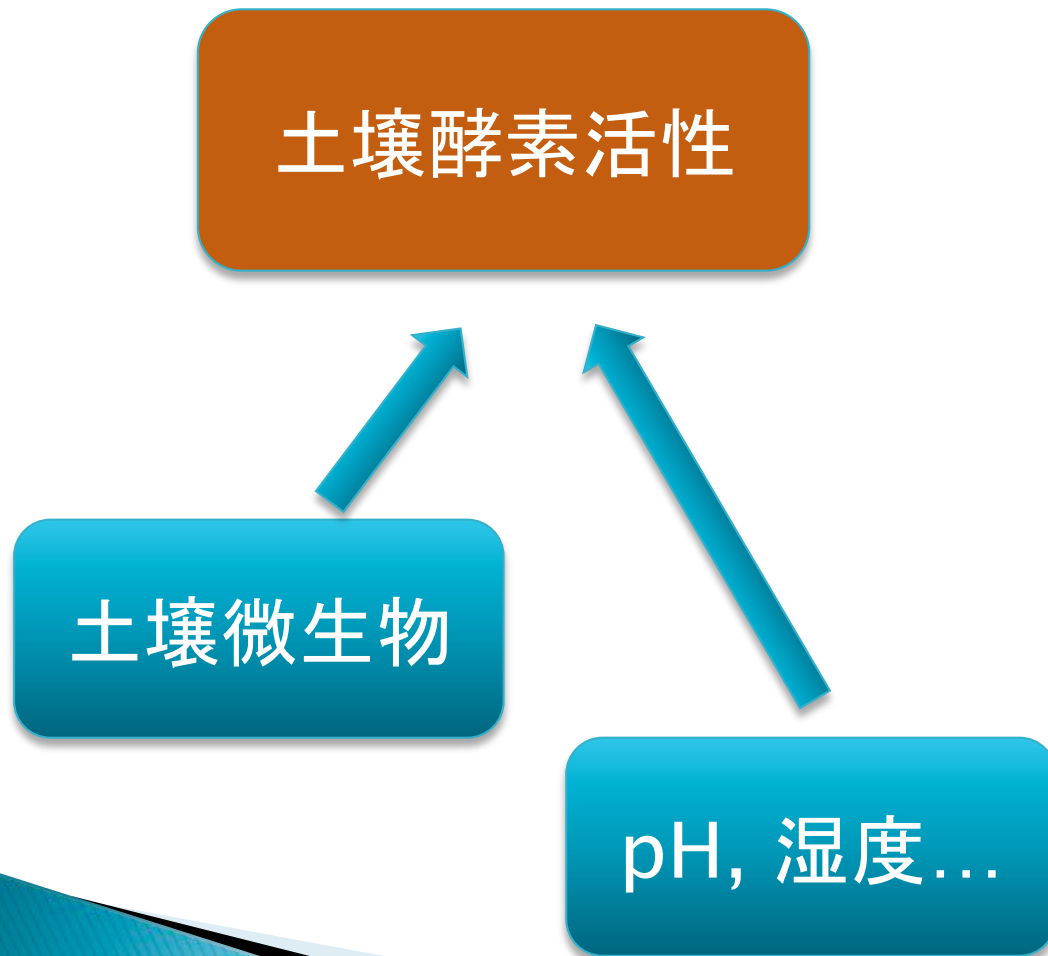
# RDA/CCA



その場合は→Partial RDA/CCA



# RでRDAを実践



- ▶ 土壌酵素活性データ。
- ▶ 微生物群集組成と物理化学特性で説明をしたい。

# RでRDAを実践

土壤酵素活性  
(4種類測定)

=

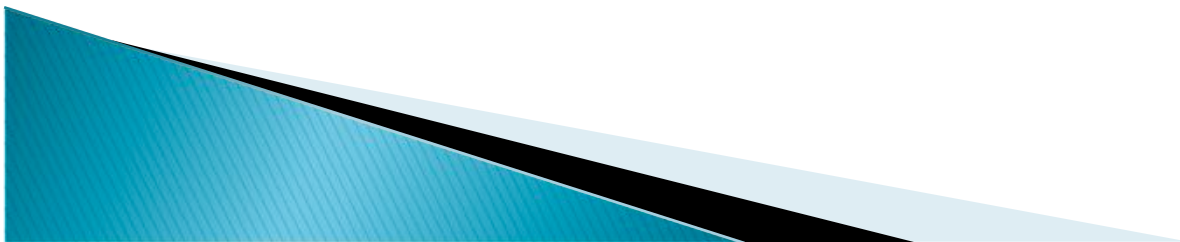
土壤微生物

+

物理化学条件

+

残差



# Variation Partitioning

*Ecology*, 87(10), 2006, pp. 2614–2625  
© 2006 by the Ecological Society of America

## VARIATION PARTITIONING OF SPECIES DATA MATRICES: ESTIMATION AND COMPARISON OF FRACTIONS

PEDRO R. PERES-NETO,<sup>1</sup> PIERRE LEGENDRE, STÉPHANE DRAY, AND DANIEL BORCARD

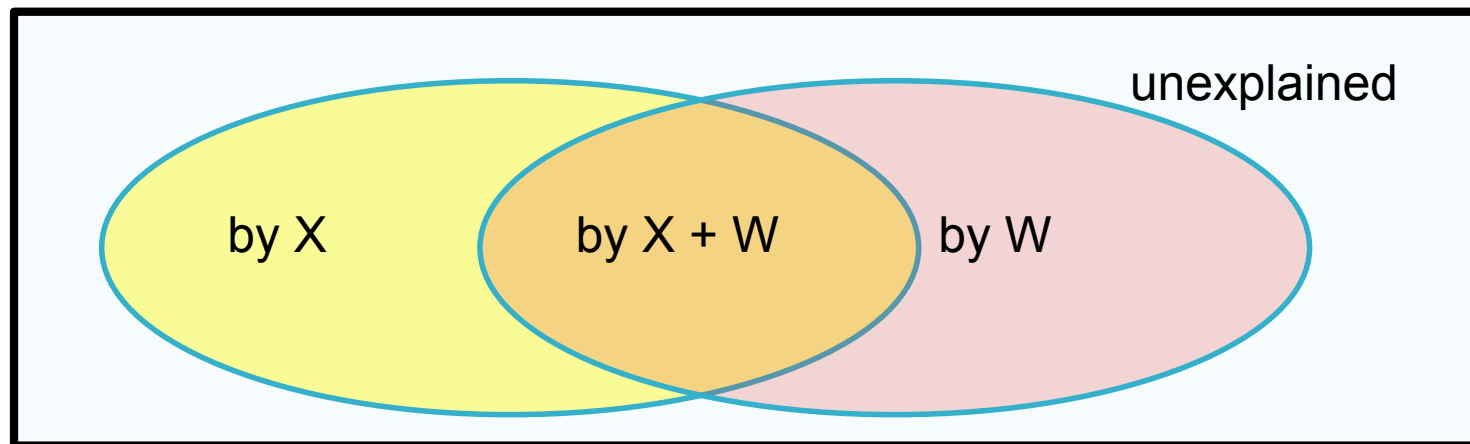
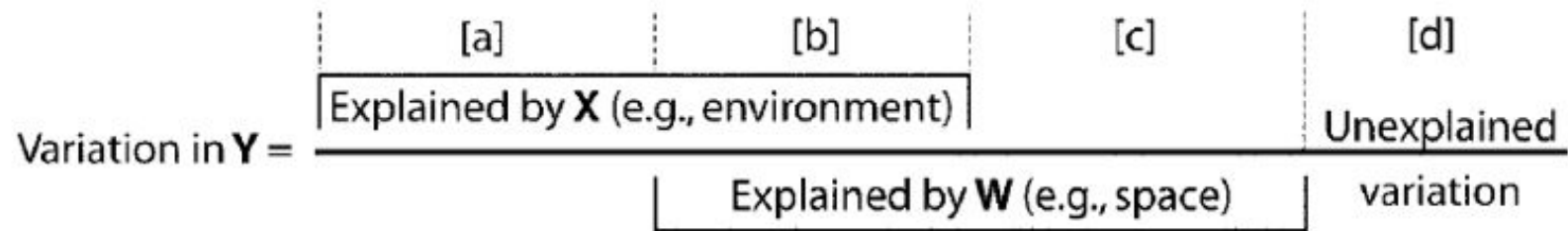
*Département des sciences biologiques, Université de Montréal, C.P. 6128, succursale Centre-ville, Montréal, Québec H3C 3J7 Canada*

*Abstract.* Establishing relationships between species distributions and environmental characteristics is a major goal in the search for forces driving species distributions. Canonical ordinations such as redundancy analysis and canonical correspondence analysis are invaluable tools for modeling communities through environmental predictors. They provide the means for conducting direct explanatory analysis in which the association among species can be studied according to their common and unique relationships with the environmental variables and other sets of predictors of interest, such as spatial variables. Variation partitioning can then be used to test and determine the likelihood of these sets of predictors in explaining patterns in community structure. Although variation partitioning in canonical analysis is routinely used in ecological analysis, no effort has been reported in the literature to consider appropriate estimators so that comparisons between fractions or, eventually, between different canonical models are meaningful. In this paper, we show that variation partitioning as currently applied in canonical analysis is biased. We present appropriate unbiased estimators. In addition, we outline a statistical test to compare fractions in canonical analysis. The question addressed by the test is whether two fractions of variation are significantly different from each other. Such assessment provides an important step toward attaining an understanding of the factors patterning community structure. The test is shown to have correct Type I error rates and good power for both redundancy analysis and canonical correspondence analysis.

*Key words:* adjusted coefficient of determination; bootstrap; canonical analysis; canonical correspondence analysis (CCA); ecological community; redundancy analysis (RDA); variation partitioning.



# Variation Partitioning





# Variation Partitioning

```
> lipid.vp
```

```
Partition of variation in RDA
```

```
Call: varpart(Y = lipid[, 8:11], X = lipid[, 16:33], lipid[, c(5, 12, 14:15)], scale = TRUE)
Columns of Y were scaled to unit variance
```

```
Explanatory tables:
```

```
X1: lipid[, 16:33]
```

```
X2: lipid[, c(5, 12, 14:15)]
```

```
No. of explanatory tables: 2
```

```
Total variation (SS): 388
```

```
Variance: 4
```

```
No. of observations: 96
```

```
Partition table:
```

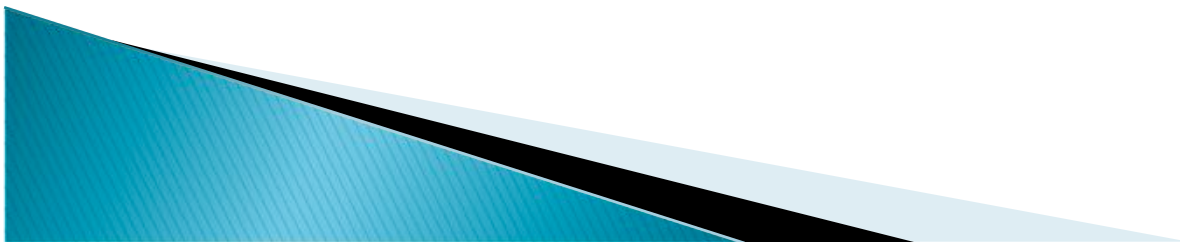
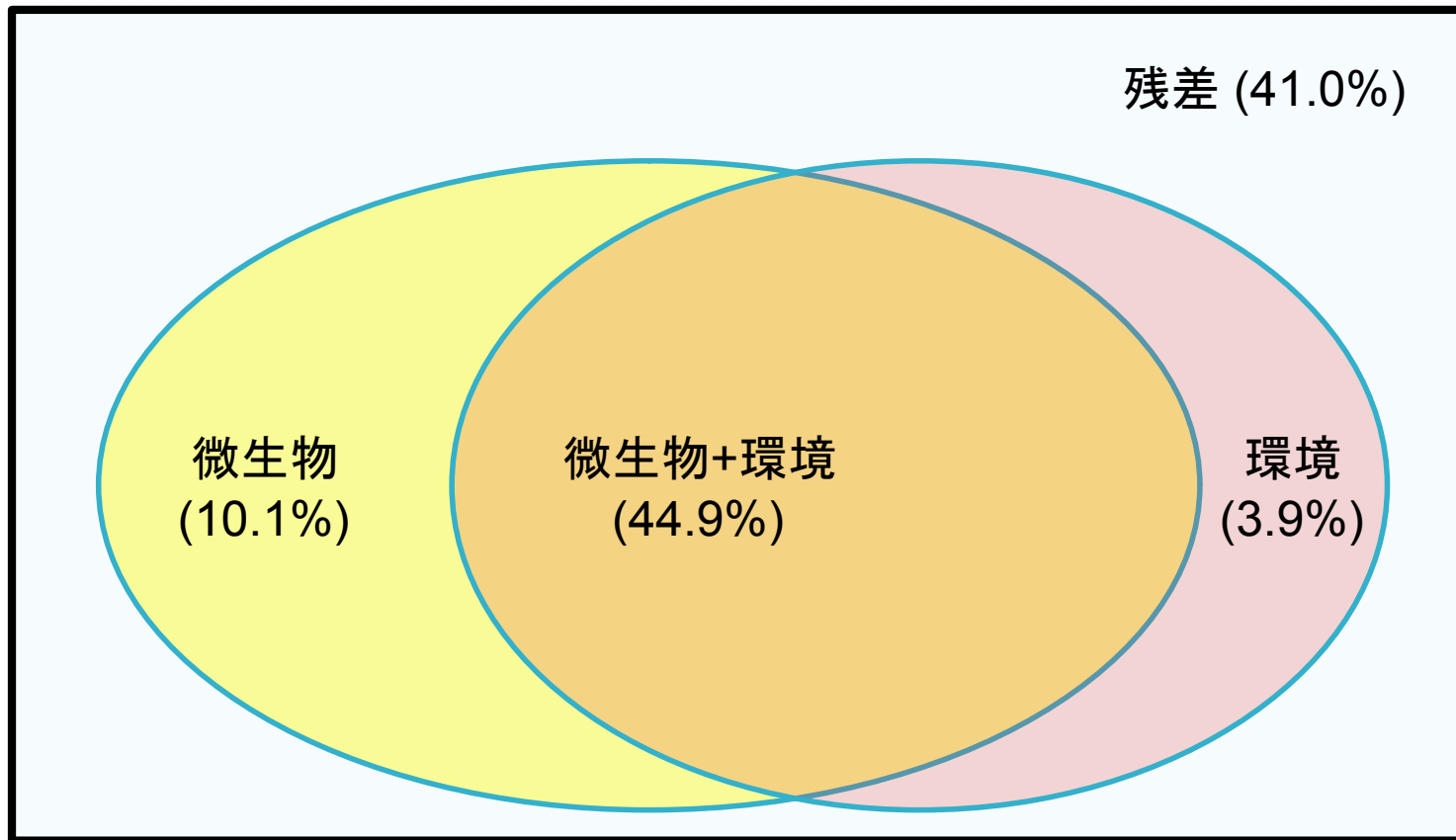
	DF	R.squared	Adj.R.squared	Testable
[a+b] - X1	18	0.63596	0.55085	TRUE
[b+c] - X2	4	0.50986	0.48832	TRUE
[a+b+c] - X1+X2	22	0.68478	0.58978	TRUE
Individual fractions				
[a] - X1 X2	18		0.18146	TRUE
[b]	0		0.44939	FALSE
[c] - X2 X1	4		0.03893	TRUE
[d] - Residuals			0.41022	FALSE

```
---
```

```
Use function 'rda' to test significance of fractions of interest
```

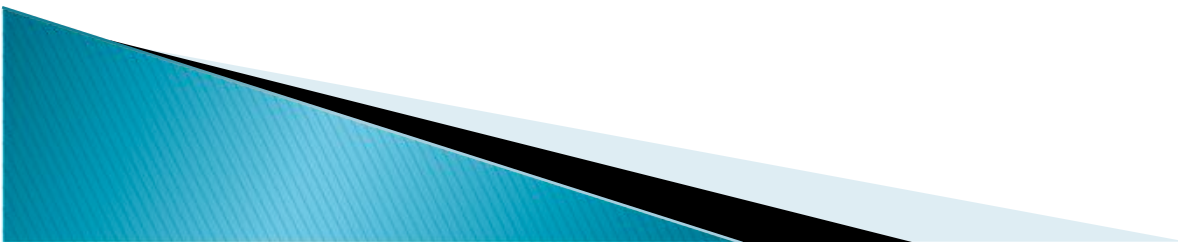
# Variation Partitioning

酵素活性の分散 (100%)



# その他の多変量解析

- ▶ Nonmetrical Multidimensional Scaling (NMS/NMDS)
- ▶ クラスター解析
- ▶ 判別分析
- ▶ 数量化理論
  
- ▶ 今回は説明無し



# どの多変量解析を使うか?

データの圧縮、要約、分類分け

ある多変量データを他の多変量データで説明したい

変数間の関係は一山型?  
データは広範囲の環境条件下から?  
DCAのgradient length > 4

変数間の関係は一山型?  
データは広範囲の環境条件下から?  
DCAのgradient length > 4

No

Yes

PCA

CA/DCA

No

Yes

RDA

CCA

NMDSを使う事も少し検討してみる

さらに詳しく分析  
したい人

Partial RDA/CCA  
Permutation test  
Variation partitioning

# 参考文献

- ▶ 土壌動物群集の研究における座標付け手法の活用.  
長谷川元洋 (2006) *Edaphologia* 80:35–64
- ▶ Variation partitioning of species data matrices: estimation and comparison of fractions. Peres-Neto et al. (2006) *Ecology* 87: 2614–2625
- ▶ Analyzing Ecological Data. Zuur et al. (2007)
- ▶ ordinationテクニックに関するHP  
<http://ordination.okstate.edu/>

