

# **PŘÍMÁ ORDINAČNÍ ANALÝZA**

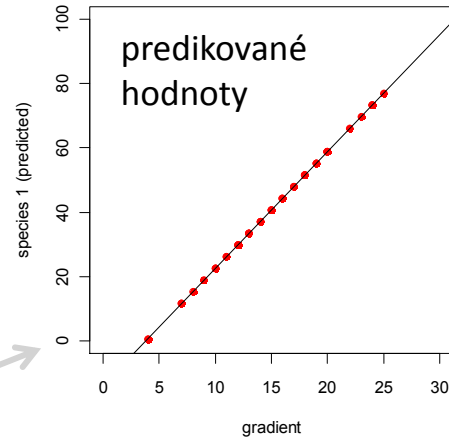
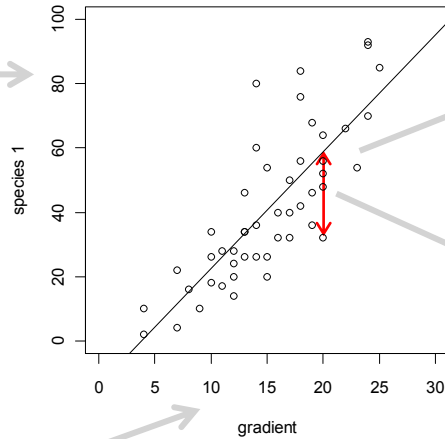
# PŘEHLED METOD ORDINAČNÍ ANALÝZY

	raw-data-based (založené na primárních datech)		distance-based (založené na distanční matici)
	linear (lineární)	unimodal (unimodální)	
unconstrained (nepřímé)	PCA (analýza hlavních komponent)	CA, DCA (korespondenční a detrendovaná korespondenční analýza)	PCoA (analýza hlavních koordinát)  NMDS (nemetrické mnohorozměrné škálování)
constrained (přímé)	RDA (redundanční analýza)	CCA (kanonická korespondenční analýza)	db-RDA (redundanční analýza založená na distanční matici)

# PRINCIP PŘÍMÉ ORDINAČNÍ ANALÝZY (RDA)

	spe 1	spe 2	spe 3
sam 1			
sam 2			
sam 3			
sam 4			
sam 5			
sam 6			
sam 7			

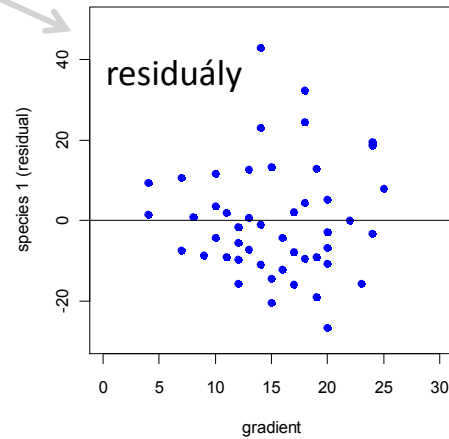
regrese abundance  
druhu na  
proměnné  
prostředí



	spe 1	spe 2	spe 3
sam 1			
sam 2			
sam 3			
sam 4			
sam 5			
sam 6			
sam 7			

	env 1	env 2
sam 1		
sam 2		
sam 3		
sam 4		
sam 5		
sam 6		
sam 7		

matice s vysvětlujícími  
proměnnými



	spe 1	spe 2	spe 3
sam 1			
sam 2			
sam 3			
sam 4			
sam 5			
sam 6			
sam 7			

## OMEZENÉ ORDINAČNÍ OSY

- Skóre založené na fitovaných hodnotách z mnohonásobné. regrese druhů (RDA), nebo vzorků (CCA) na proměnných prostředí
- Ordinační osy jsou lineárními kombinacemi prediktorů
- Reziduální variabilita je vyjádřena neomezenými osami, které jsou přítomné v každé přímé ordinaci.

# PŘÍMÁ ORDINACE

## INTERPRETACE VÝSLEDKŮ

RDA

	Inertia	Proportion	Rank
Total	0.3783	1.0000	
Constrained	0.1063	0.2808	2
Unconstrained	0.2721	0.7192	24

Inertia is variance

$R^2 = 28.08 \%$

Eigenvalues for constrained axes:

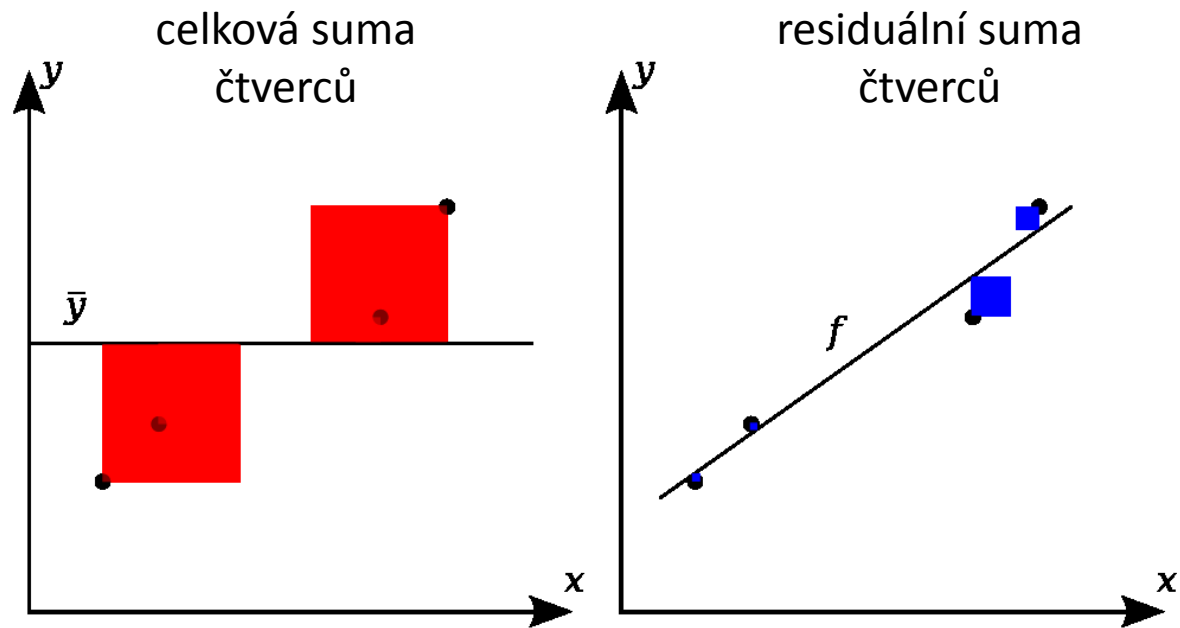
RDA1	RDA2
0.09240	0.01385

Eigenvalues for unconstrained axes:

PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
0.09543	0.03586	0.02688	0.02280	0.01706	0.01550	0.00999	0.00880

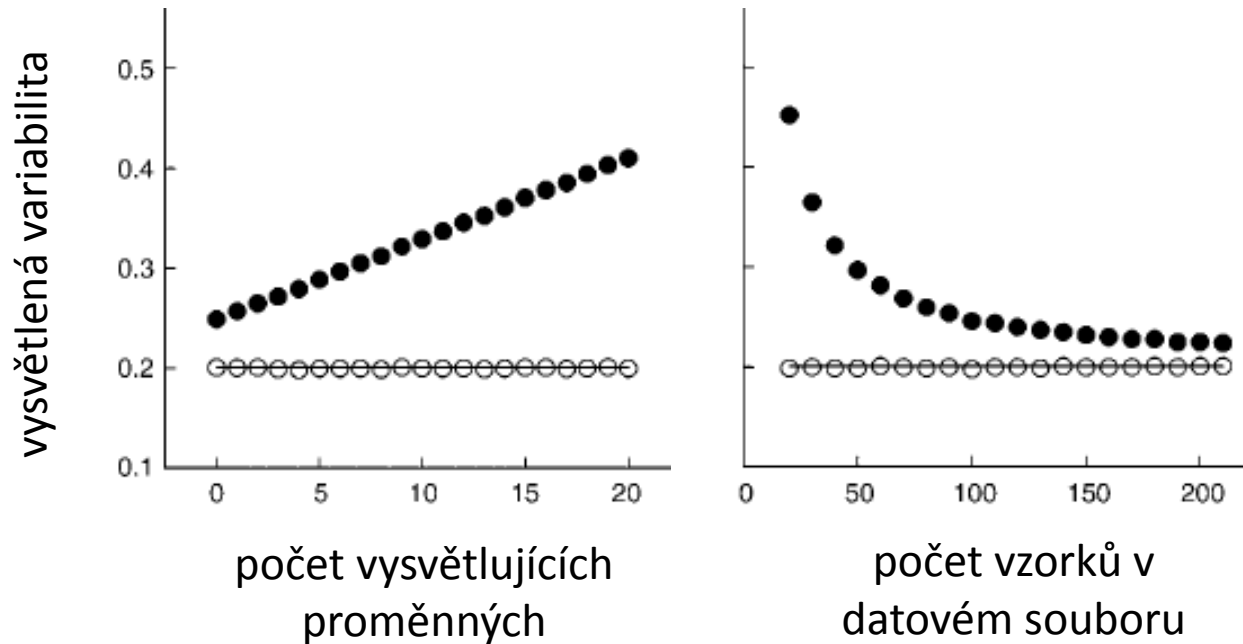
(Shown only 8 of all 24 unconstrained eigenvalues)

# KOEFICIENT DETERMINACE V REGRESI



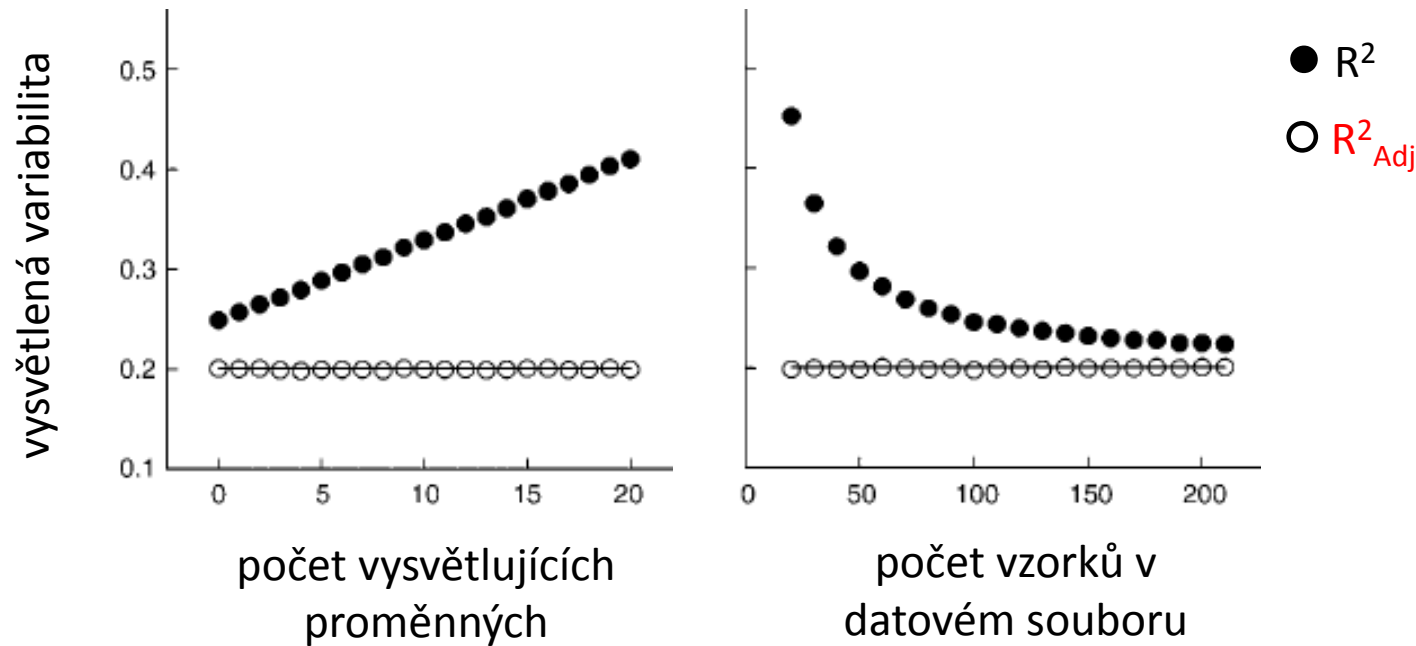
$$R^2 = 1 - \frac{SS_{\text{res}}}{SS_{\text{tot}}}$$

# VYSVĚTLENÁ VARIABILITA ( $R^2$ )



- vysvětlená variabilita stoupá s počtem vysvětlujících proměnných (i když jsou náhodné) a klesá s počtem vzorků v datovém souboru
- platí pro přímou (kanonickou) ordinační analýzu i mnohonásobnou regresi

# VYSVĚTLENÁ VARIABILITA ( $R^2$ ) A UPRAVENÝ (ADJUSTED) $R^2$



- upravený  $R^2$  se nemění s počtem vysvětlujících proměnných a počtem vzorků v souboru



# VÝPOČET UPRAVENÉHO $R^2$ POMOCÍ EZEKIELOVY FORMULE (RDA)

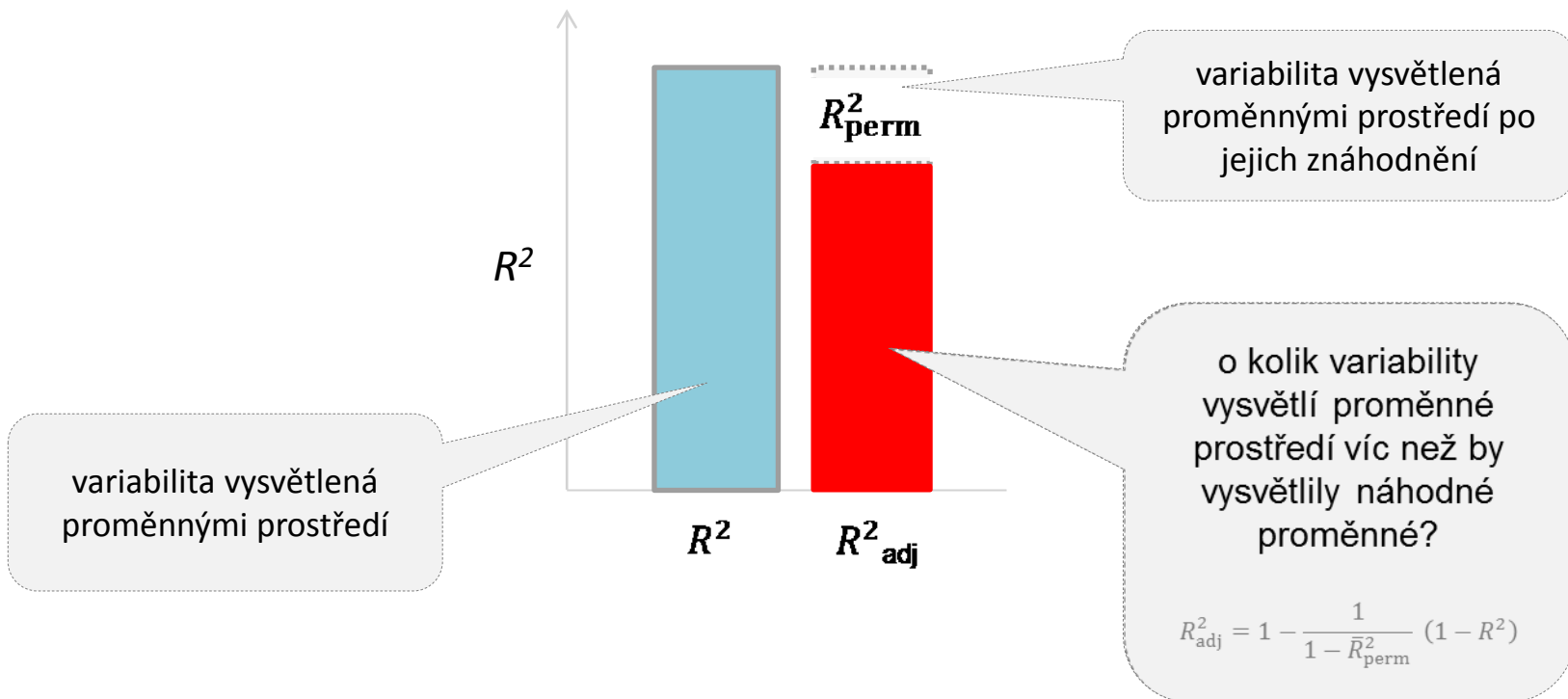
$$R^2_{(Y|X)adj} = 1 - \frac{n-1}{n-p-1} (1 - R^2_{Y|X})$$

$n$  ... počet vzorků

$p$  ... počet vysvětlujících proměnných

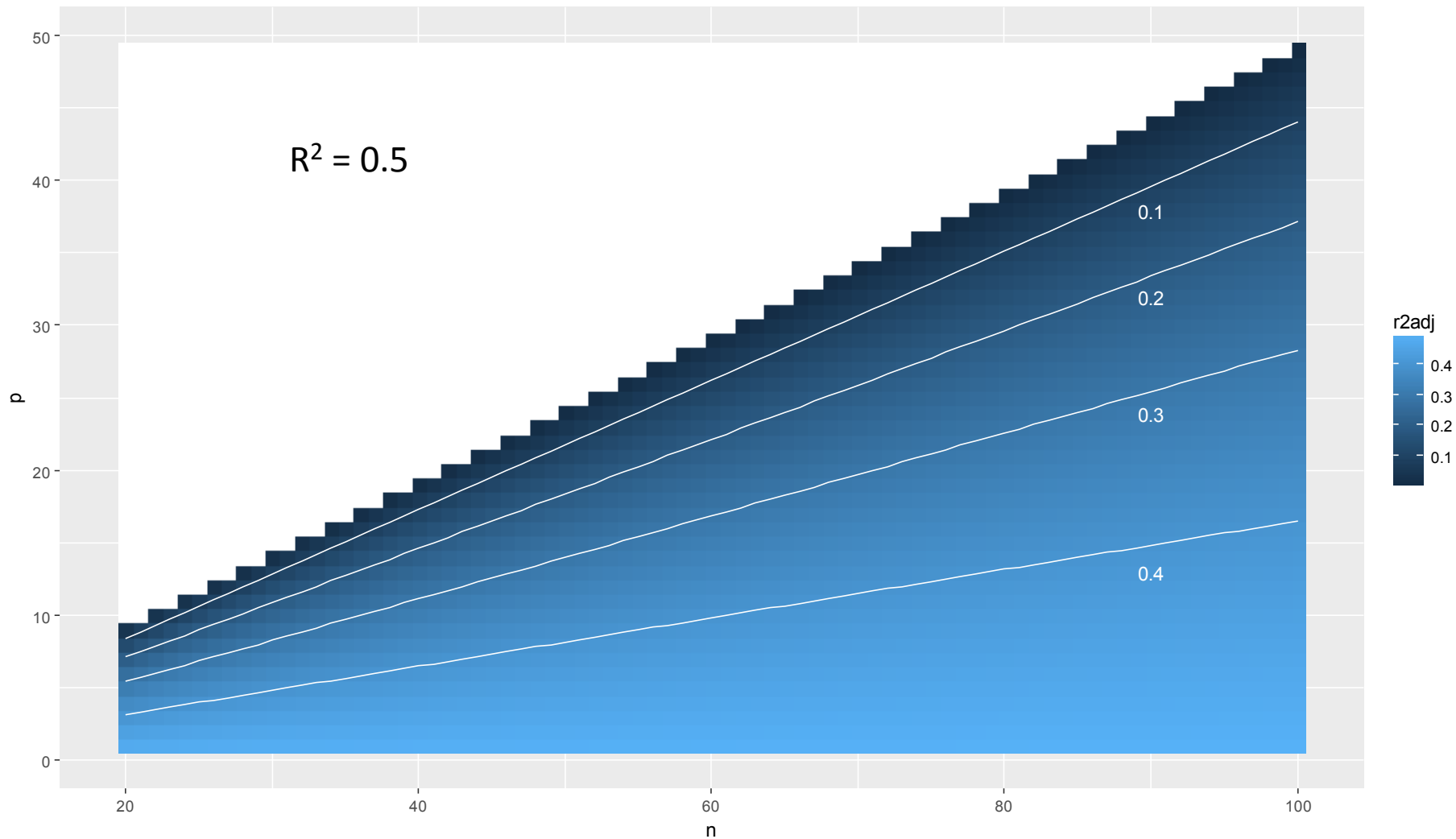
$R^2_{Y|X}$  ... vysvětlená variabilita bez adjustace

## Výpočet adjustovaného $R^2$ permutačním modelem (RDA, CCA)



# UPRAVENÉ $R^2$ PRO KOMBINACE N A P

$$R^2_{(Y|X)\text{adj}} = 1 - \frac{n-1}{n-p-1} (1 - R^2_{Y|X})$$



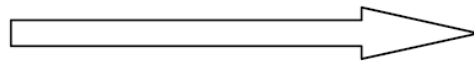
# PŘÍMÁ ORDINAČNÍ ANALÝZA

## MONTE-CARLO PERMUTAČNÍ TEST

- testuje nulovou hypotézu, že druhové složení je nezávislé na jedné nebo více vysvětlujících proměnných pomocí randomizace

ošetřeno	plocha	pokr.
1	1	8
1	2	7
1	3	7
1	7	6
1	8	5
1	9	6
0	4	5
0	5	5
0	6	6
0	10	1
0	11	2
0	12	2
	F =	0.413

RANDOMIZACE: Každé ploše se přiřadí náhodně jeden vegetační zápis (v našem případě jedna hodnota pokravnosti). Tím se zruší jakákoliv závislost mezi proměnnou prostředí (tj. ošetřením) a hodnotami pokravnosti.



Z hodnot pokravnosti spočtu množství variability vysvětlené proměnnými prostředí (F-statistika). Pro hodnoty:

výchozí                      randomizované

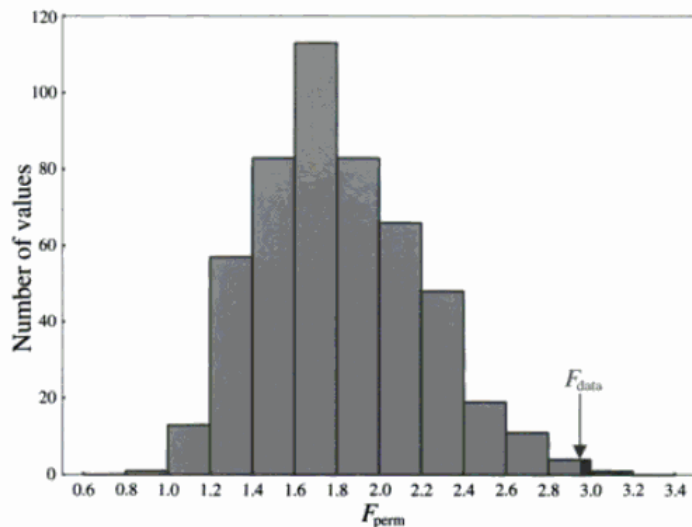
ran.1	ran.2	ran.3	ran.4
7	2	5	7
6	5	8	2
5	6	1	5
1	8	6	7
6	1	7	6
6	5	6	2
8	5	6	5
5	6	7	1
5	7	2	6
2	6	2	5
7	2	5	6
2	7	5	8
0.175	0.005	0.005	0.076

A tak dále.  
Až po dosažení kýženého počtu randomizací.

# PŘÍMÁ ORDINAČNÍ ANALÝZA

## MONTE-CARLO PERMUTAČNÍ TEST

- Randomizace produkuje rozdělení hodnot, které by mělo testové kritérium za platnosti nulové hypotézy
- test první kanonické osy – vliv jen jedné kvantitativní proměnné
- test všech kanonických os – vliv všech proměnných, nebo vliv jedné kategoriální proměnné s více kategoriemi (počet os = počet kategorií – 1)
- testová statistika –  $F_{\text{data}}$  (**pseudo-F**)



$$P = \frac{n_x + 1}{N + 1}$$

P – hladina signifikance

$n_x$  – počet permutací, kde  $F_{\text{perm}} \geq F_{\text{data}}$

N – celkový počet permutací

# POSTUPNÝ VÝBĚR VYSVĚTLUJÍCÍCH PROMĚNNÝCH

## *FORWARD SELECTION*

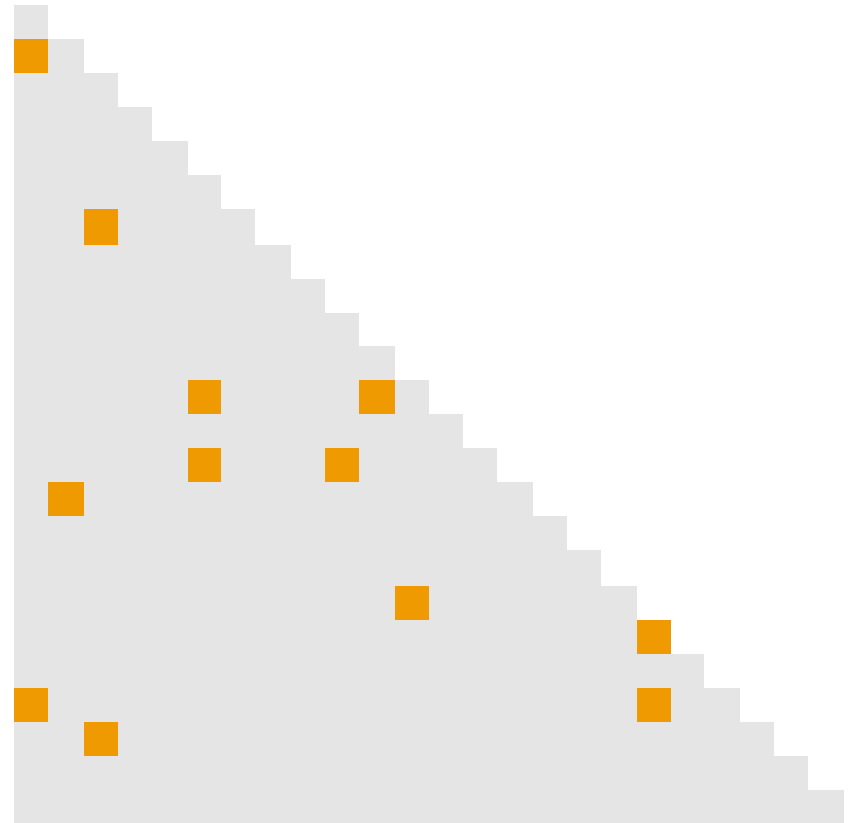
- ze souboru vysvětlujících proměnných umožňuje vybrat jen ty, které mají průkazný vliv
- v každém kroku testuje zvlášť vliv jednotlivých proměnných (Monte-Carlo permutační test)
- vybere tu proměnnou, která vysvětlí nejvíce variability a zároveň je signifikantní; tuto proměnnou pak do modelu zahrne jako kovariátu
- v dalším kroku znovu testuje vliv jednotlivých proměnných na druhová data (s odstraněním vlivu kovariát) a opakuje předchozí kroky
- testy signifikance jsou zatíženy mnohonásobným porovnáním, a jsou proto poměrně liberální (počet signifikantních proměnných je často nerealisticky vysoký a vyžaduje např. Holmovu korekci)
- Obrácený postup – backward selection
  - Nejprve se fituje saturovaný model se všemi prediktory a případnými interakcemi
  - Postupně se z něj odstraňují neprůkazné proměnné s nejmenším vlivem
  - Speciálně vhodný pro experimentální data

# PROBLÉM MNOHONÁSOBNÉHO POROVNÁNÍ

## FORWARD SELECTION PŘÍLIŠ LIBERÁLNÍ

### Simulace:

- 25 náhodně vygenerovaných proměnných
- otestování průkaznosti korelace každé proměnné s každou (čtvercová matice)
- průkazné korelace ( $p < 0.05$ ) jsou označeny červeně
- dohromady 300 analýz, z nich je 16 průkazných



# ŘEŠENÍ PŘÍLIŠ LIBERÁLNÍ FORWARD SELECTION ALE PŘÍLIŠ RESTRIKTIVNÍCH KOREKCÍ NA MNOHONÁSOBNÉ TESTOVÁNÍ

- globální test
- pokud významný, pak forward selection
  - Výhradně směr forward
- končí, pokud:
  - $R^2_{Adj}$  má překročit globální  $R^2_{Adj}$
  - by přidávaná proměnná nebyla významná

Blanchet et al. 2008

<https://esajournals.onlinelibrary.wiley.com/doi/abs/10.1890/07-0986.1>

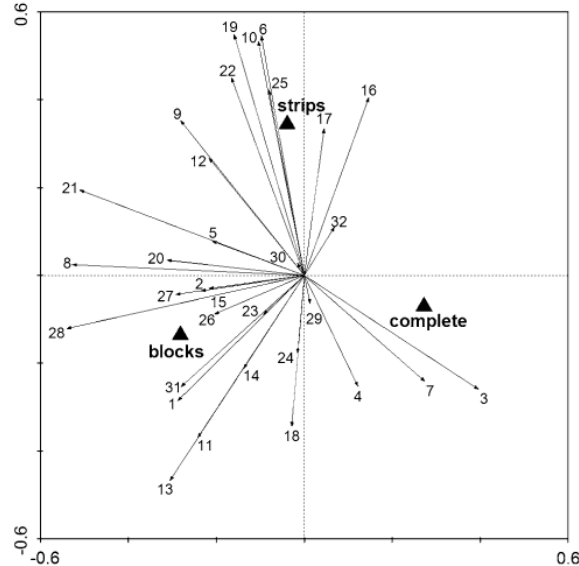
# PERMUTAČNÍ TESTY MOHOU TESTOVAT

- Celkový model (=všechny omezené ordinační osy)
- Jednotlivé omezené ordinační osy (typicky 1.)
- Jednotlivé prediktory
  - Jednoduché efekty (simple effects ve veganu; marginal effects v Canoco terminologii)
    - Ignorují efekty ostatních prediktorů
    - Odpovídají jednoduchým párovým korelacím mezi maticí odpovědi a prediktorem
    - Typicky slouží jako podklad pro kreslení ordinačního diagramu
    - Nezáleží na pořadí prediktorů
  - Parciální efekty (partial effects – v Canoco terminologii, obecně rozšířeno; marginal effects ve veganu – pozor na to!)
    - Uvažují vliv ostatních proměnných v modelu
    - Odpovídají unikátní vysvětlovací síle prediktoru vzhledem k ostatním prediktorům
    - Nezáleží na pořadí prediktorů
  - Sekvenční efekty
    - Testy prediktorů v modelu v sekvenčním pořadí
    - Berou v potaz efekty prediktorů uvedených před daným prediktorem, ignorují následující prediktory
    - Záleží na pořadí, ale mají smysl jako charakteristika modelu pokud jsou prediktory řazeny od toho, co vysvětluje nejvíce po ten co vysvětluje nejméně

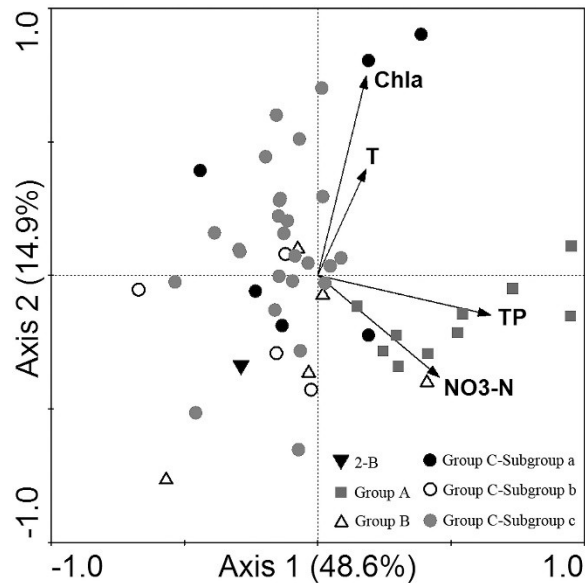


# ORDINAŘNÍ DIAGRAM PŘÍMÉ ORDINACE

- Kontinuální prediktory – šipky
- Kategořální prediktory – centroidy
- Nezáleží na typu ordinance (unimodální/lineární/distance-based)



**Fig. 3** Ordination diagram (RDA) showing association of butterflies with the three hay mowing regimes. 1, *Aglais urticae*; 2, *Aphantopus hyperanthus*; 3, *Araschnia levana*; 4, *Argynnis paphia*; 5, *Aricia agestis*; 6, *Boloria dia*; 7, *Celastrina argiolus*; 8, *Coenonympha pamphilus*; 9, *Colias hyale*; 10, *Erynnis tages*; 11, *Gonepteryx rhamni*; 12, *Inachis io*; 13, *Issoria lathonia*; 14, *Lasiommata megera*; 15, *Leptidea reali*; 16, *Lycaena phlaeas*; 17, *Lycaena tityrus*; 18, *Phengaris nausithous*; 19, *Phengaris teleius*; 20, *Maniola jurtina*; 21, *Melanargia galathea*; 22, *Ochlodes sylvanus*; 23, *Papilio machaon*; 24, *Pieris brassicae*; 25, *Pieris napi*; 26, *Pieris rapae*; 27, *Polygonia c-album*; 28, *Polyommatus icarus*; 29, *Pontia daplidice*; 30, *Thymelicus lineola*; 31, *Vanessa atalanta*; 32, *Vanessa cardui*



The RDA ordination plot shows the relationship between ciliate communities and significant environmental variables. The result was based on square root-transformed biotic and log-transformed abiotic data of 51 effective samples in the northern Beibu Gulf in August 2011. T-water temperature; Chl  $\alpha$ - chlorophyll- $\alpha$ ; Sal-salinity; 2-B = bottom sample of station 2.

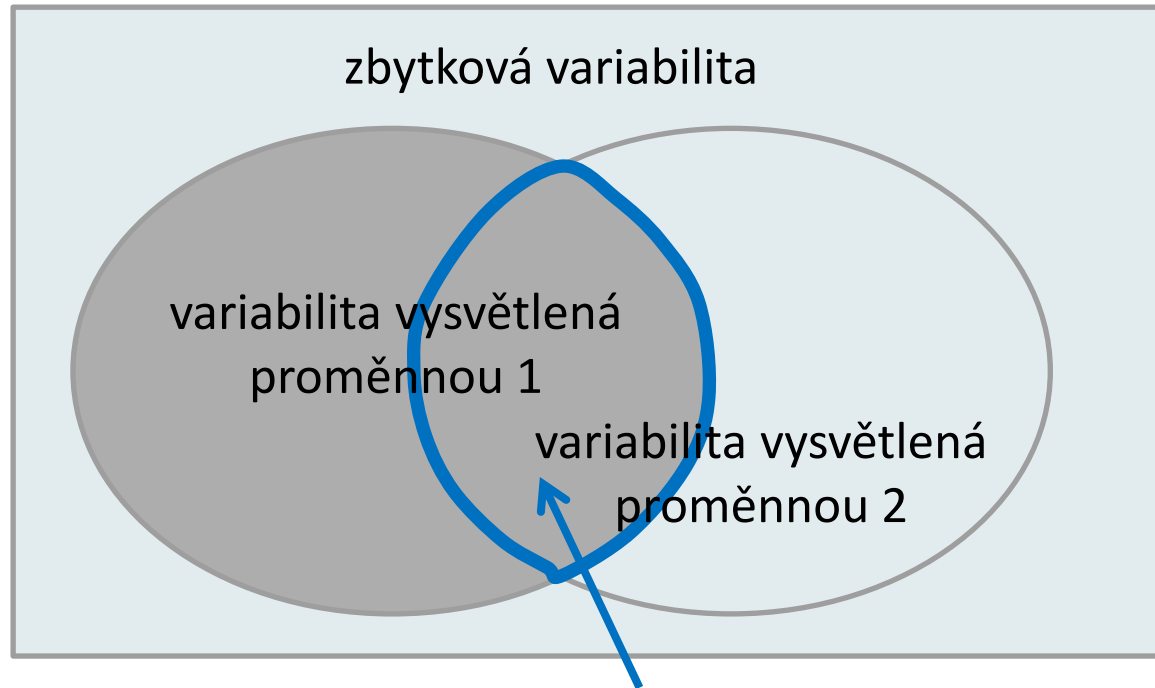
# PARCIÁLNÍ ORDINACE

## *PARTIAL ORDINATION*

- odstraňuje část variability vysvětlené proměnnými, jejichž vliv chceme odečíst, ale ne přímo kvantifikovat (například vliv umístění ploch do bloků)
- „odečítané“ proměnné se definují jako **kovariáty**
- pokud následuje přímá ordinace – ordinační osy představují čistý vliv ostatních vysvětlujících proměnných bez vlivu kovariát
- pokud následuje nepřímá ordinace – ordinační osy zachycují zbytkovou variabilitu v druhových datech po odstranění vlivu kovariát
- **Pomocí kovariát testujeme i parciální (čisté) efekty více prediktorů, pokud jsou tyto navzájem korelované**
- **Kovariáty slouží ke specifikaci designů experimentů (nebo pozorování) v případě strukturovaných dat**

# ROZKLAD VARIABILITY

## VARIATION PARTITIONING



**vysvětlená variabilita sdílená proměnnou 1 a proměnnou 2 kvůli jejich vzájemné korelaci**

- Lze aplikovat i na skupiny proměnných
- Lze „naslepo“ použít i s neprůkaznými proměnnými (odpadá problém s postupným výběrem)

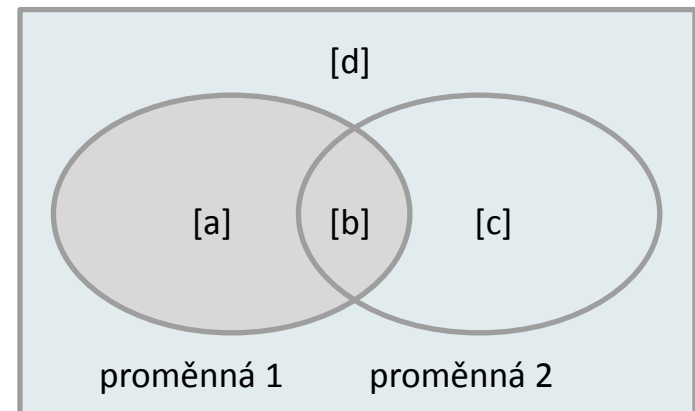
Borcard et al. 1992, *Ecology* 73: 1045–1055

# ROZKLAD VARIABILITY

## VARIATION PARTITIONING

Tři analýzy:

vysvětlující proměnná	kovariáta	vysvětlená variabilita
1 a 2	není	$[a]+[b]+[c]$
1	2	$[a]$
2	1	$[c]$



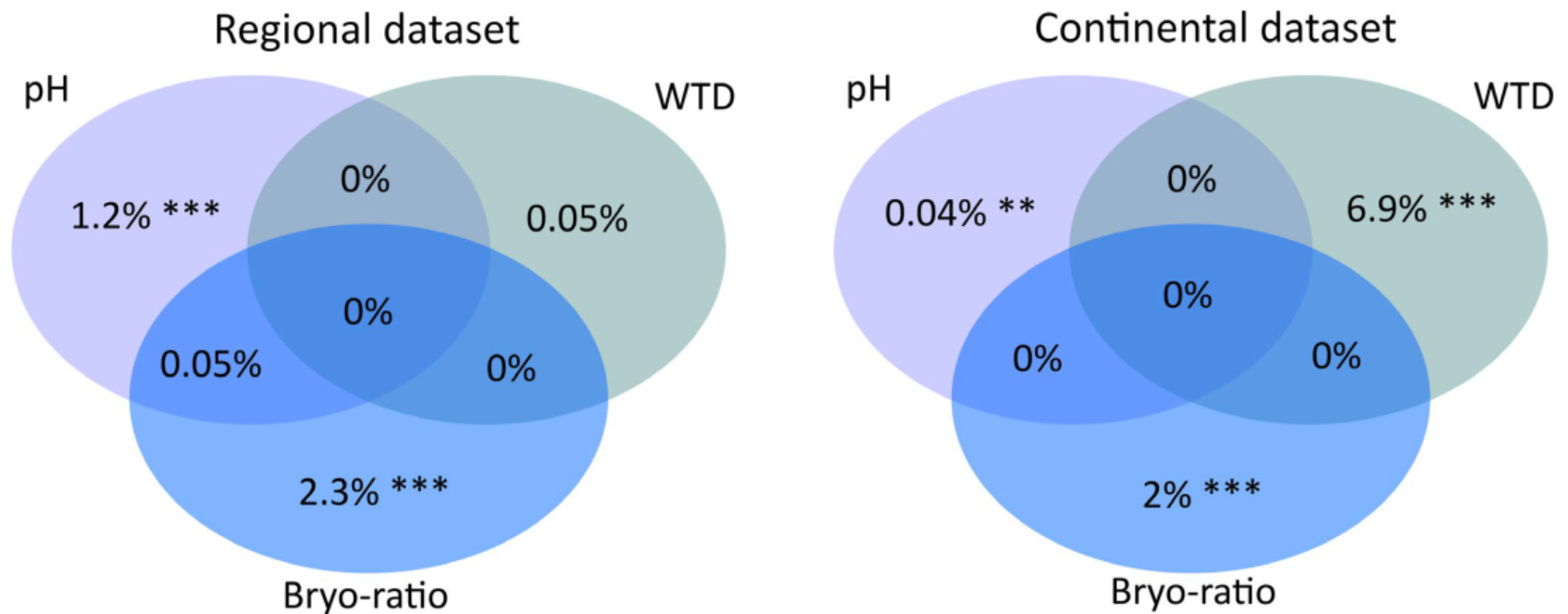
sdílená variabilita  $[b] = ([a]+[b]+[c]) - [a] - [c]$

nevysvětlená variabilita  $[d] = Total\ inertia - ([a]+[b]+[c])$

$[a]+[b]$  – celkový (*marginal*) vliv proměnné 1

$[a]$  – čistý (*partial, conditional*) vliv proměnné 1 (bez vlivu prom. 2)

# VARIATION PARTITIONING PRO VÍCE SKUPIN PROMĚNNÝCH



**FIGURE 3** Individual contributions of *pH*, *WTD* and *bryo-ratio* in the compositional variation of *target species* (the rich fen-specialized vascular plant community) in the regional and continental data sets. *p*-Values are indicated with asterisks (\*  $p < 0.05$ ; \*\*  $p < 0.01$ ; \*\*\*  $p < 0.001$ ) [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]