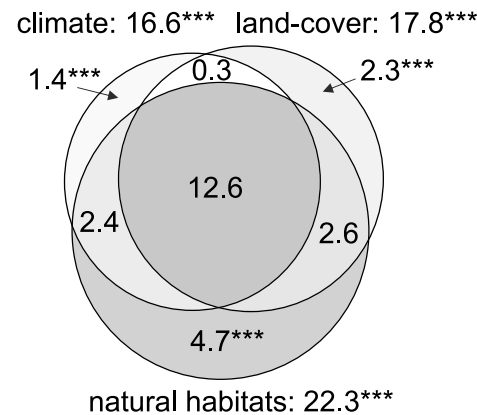
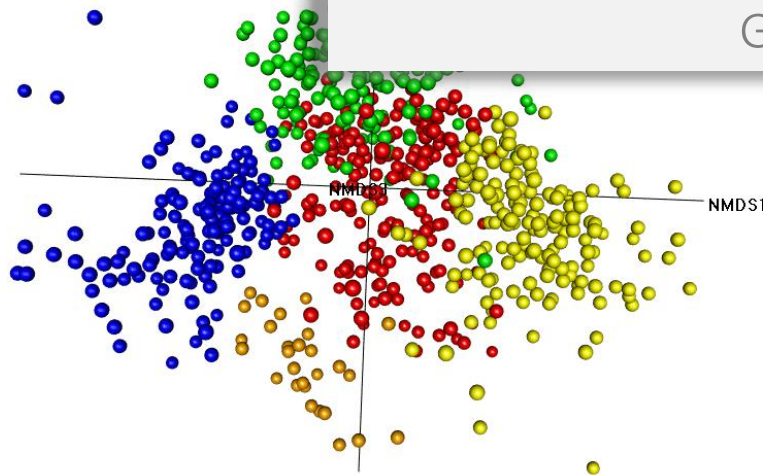


Metody fyzické geografie 3: Biogeografie & ekologie

Jan Divíšek
Geografický ústav & Ústav botaniky a zoologie



I TÝ SI ZAPIŠ NOVÝ
PŘEDMĚT Z 8055
METODY FYZICKÉ
GEOGRAFIE 3!



Metody fyzické geografie 3 – 31. 10. 2017

- Teoretická část
 - Forward selection pro regresi
 - Variation partitioning pro regresi
- Praktická část
 - Lineární regrese + forward selection a variation partitioning

Forward selection a variation partitioning v lineární regresí

Lineární regresní model v R

```
m <- lm(Y ~ X1 + X2 + ... + X5)
```

```
anova(m)
```

Analysis of Variance Table

Response: Species

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
Altitude	1	503.89	503.89	31.3657	6.541e-05	***
Slope	1	2.61	2.61	0.1622	0.693243	
pH	1	182.77	182.77	11.3768	0.004551	**
Moisture	1	76.63	76.63	4.7702	0.046465	*
E3_cover	1	31.73	31.73	1.9753	0.181690	
Residuals	14	224.91	16.07			

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Lineární regresní model v R

```
m <- lm(Y ~ X1 + X2 + ... + X5)
```

```
anova(m)
```

Analysis of Variance Table

Response: Species

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
pH	1	667.23	667.23	41.5325	1.536e-05	***
Slope	1	6.50	6.50	0.4044	0.53511	
Altitude	1	15.55	15.55	0.9678	0.34192	
Moisture	1	76.63	76.63	4.7702	0.04647	*
E3_cover	1	31.73	31.73	1.9753	0.18169	
Residuals	14	224.91	16.07			

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

	Slope	pH	Moisture	E3_cover
Altitude	0.296	-0.759	0.268	-0.331
Slope		-0.221	0.085	-0.408
pH			-0.229	0.461
Moisture				0.149

Forward selection

- Metoda pro výběr souboru „nejlepších“ vysvětlujících proměnných z celého setu proměnných, které mám k dispozici
- Cílem je redukovat počet proměnných, ale zachovat maximální vysvětlenou variabilitu
- Dobře použitelné pro ekologické studie s korelovanými proměnnými (nikoliv pro laboratorní experimenty s propracovaným designem)
- Použitelné v lineární regresi a vícerozměrných metodách (RDA, CCA)
- V R několik funkcí
 - `ordistep {vegan}`
 - `ordiR2step {vegan}`
 - `forward.sel {packfor}`

Jak pracuje forward selection?

- Předem je nutné otestovat signifikanci celého modelu, tj. se všemi vysvětlujícími proměnnými → pokud není signifikantní, nemá smysl dělat FS
- Kroky forward selection:
 1. Každá vysvětlující proměnná se použije v samostatném modelu → zaznamená se vysvětlená variabilita
 2. Seřadí proměnné podle vysvětlené variability od „nejlepší“ po „nejhorší“
 3. Zjistí zda variabilita vysvětlená nejlepší proměnnou je statisticky signifikantní (v regresi použije F-test), pokud není → zastaví výběr
 4. Zjistí kolik variability vysvětlí každá ze zbylých proměnných zatímco první vybraná proměnná je zahrnuta v modelu jako kovariáta
 5. Seřadí proměnné podle vysvětlené variability a pro nejlepší proměnnou otestuje statistickou významnost jejího příspěvku do modelu, pokud nevýznamný → zastaví výběr
 6. Opakuje body 4 a 5 dokud další proměnné významně přispívají do modelu

Jak pracuje forward selection?

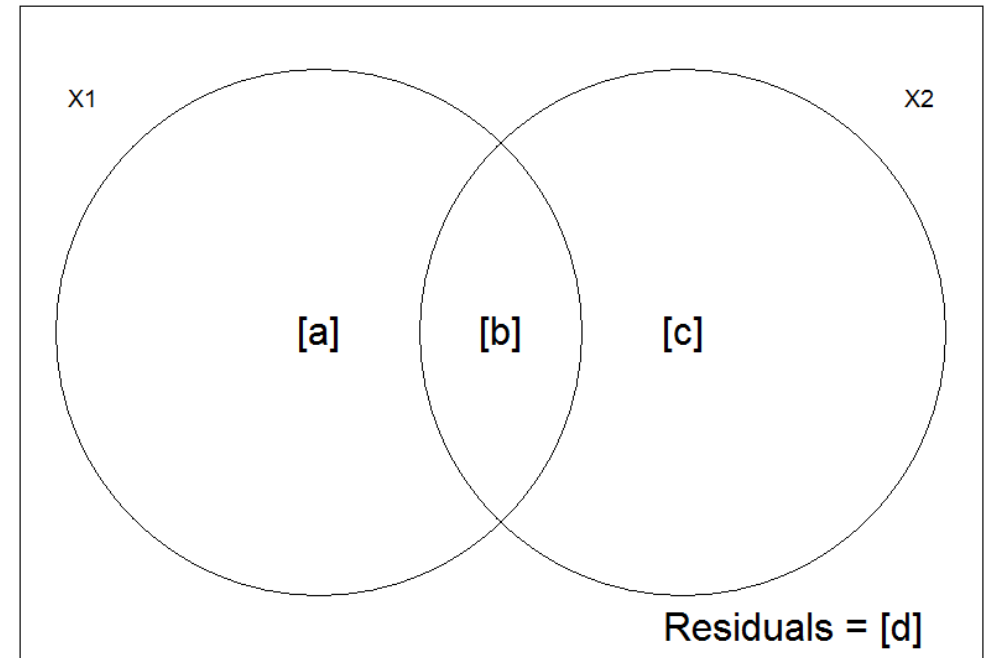
- Kritéria pro zastavení výběru
 1. Statistická signifikance
 2. Adjustovaný R^2 globálního modelu (tj. modelu se všemi proměnnými)
- Lze použít v lineární regresi a přímé ordinaci (RDA, CCA)
- Alternativy k forward selection
 - backward selection
 - forward-backward selection

Rozklad variance (*variation partitioning*)



`varpart {vegan}`

- Umožňuje rozložit variabilitu vysvětlenou danými proměnnými na následující části:
 - [a] Variabilitu vysvětlenou čistým vlivem první proměnné (nebo sadou proměnných)
 - [b] Variabilitu vysvětlenou sdíleným vlivem první a druhé proměnné (případně první a druhou sadou proměnných)
 - [c] Variabilitu vysvětlenou čistým vlivem druhé proměnné (nebo sadou proměnných)
- Je možné použít i více proměnných (jejich sad), ale většinou se končí u 3 až 4
- Lze testovat statistickou signifikanci „čistých vlivů“
- Pokud se skupiny liší počtem proměnných → adjustovaný R^2
- Čím více jsou proměnné korelované tím větší bude sdílená variabilita



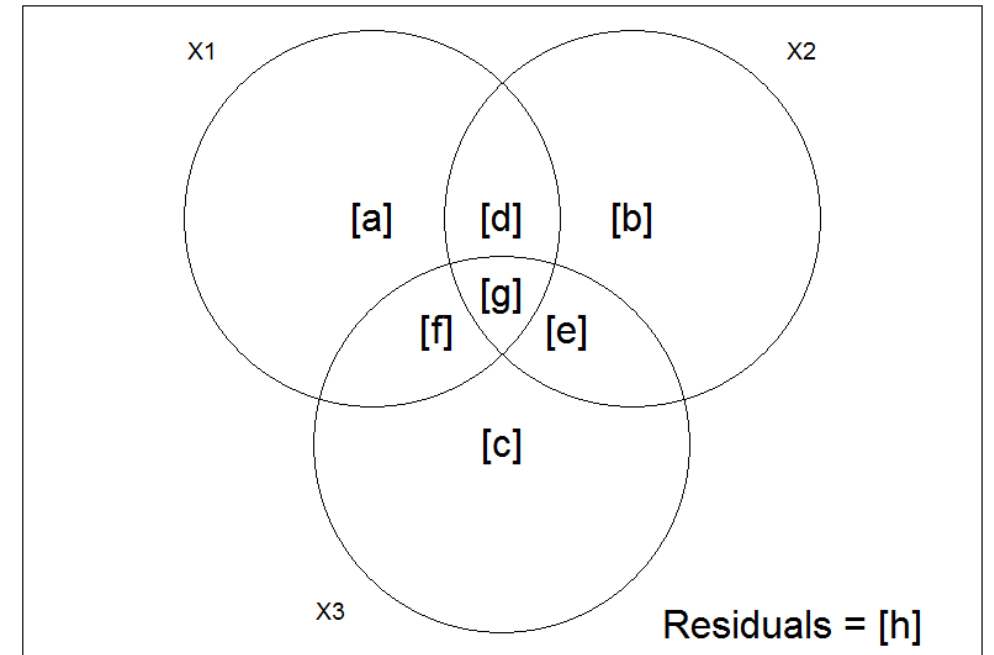
Borcard, et al. (1992)

Rozklad variance (*variation partitioning*)



varpart {vegan}

- Umožňuje rozložit variabilitu vysvětlenou danými proměnnými na následující části:
 - [a] Variabilitu vysvětlenou čistým vlivem první proměnné (nebo sadou proměnných)
 - [b] Variabilitu vysvětlenou sdíleným vlivem první a druhé proměnné (případně první a druhou sadou proměnných)
 - [c] Variabilitu vysvětlenou čistým vlivem druhé proměnné (nebo sadou proměnných)
- Je možné použít i více proměnných (jejich sad), ale většinou se končí u 3 až 4
- Lze testovat statistickou signifikanci „čistých vlivů“
- Pokud se skupiny liší počtem proměnných → adjustovaný R^2
- Čím více jsou proměnné korelované tím větší bude sdílená variabilita



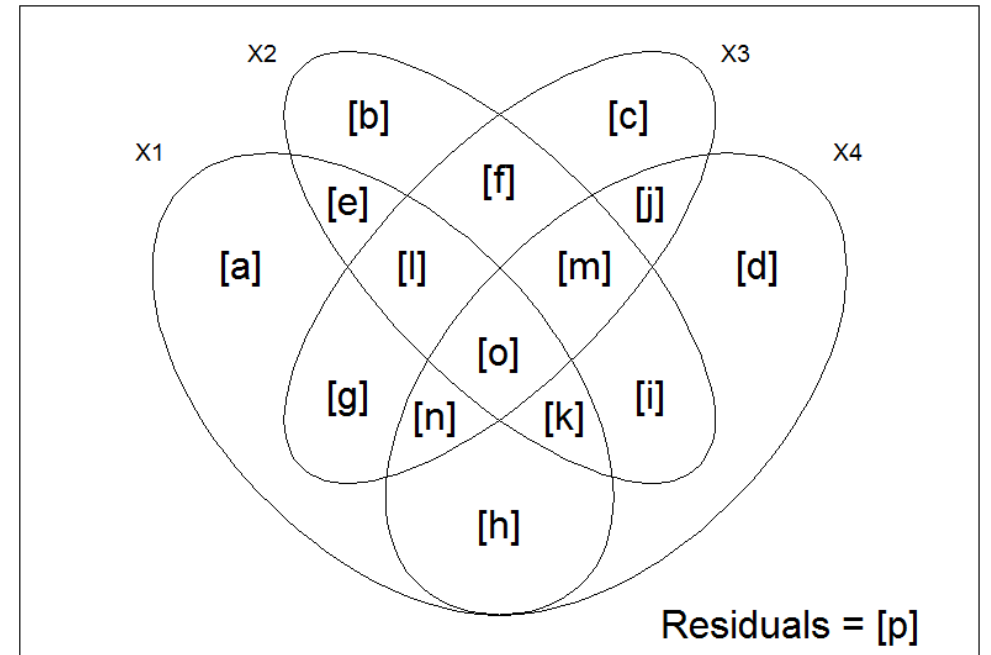
Borcard, et al. (1992)

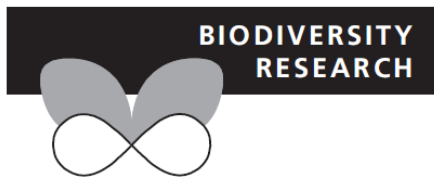
Rozklad variance (*variation partitioning*)



varpart {vegan}

- Umožňuje rozložit variabilitu vysvětlenou danými proměnnými na následující části:
 - [a] Variabilitu vysvětlenou čistým vlivem první proměnné (nebo sadou proměnných)
 - [b] Variabilitu vysvětlenou sdíleným vlivem první a druhé proměnné (případně první a druhou sadou proměnných)
 - [c] Variabilitu vysvětlenou čistým vlivem druhé proměnné (nebo sadou proměnných)
- Je možné použít i více proměnných (jejich sad), ale většinou se končí u 3 až 4
- Lze testovat statistickou signifikanci „čistých vlivů“
- Pokud se skupiny liší počtem proměnných → adjustovaný R^2
- Čím více jsou proměnné korelované tím větší bude sdílená variabilita





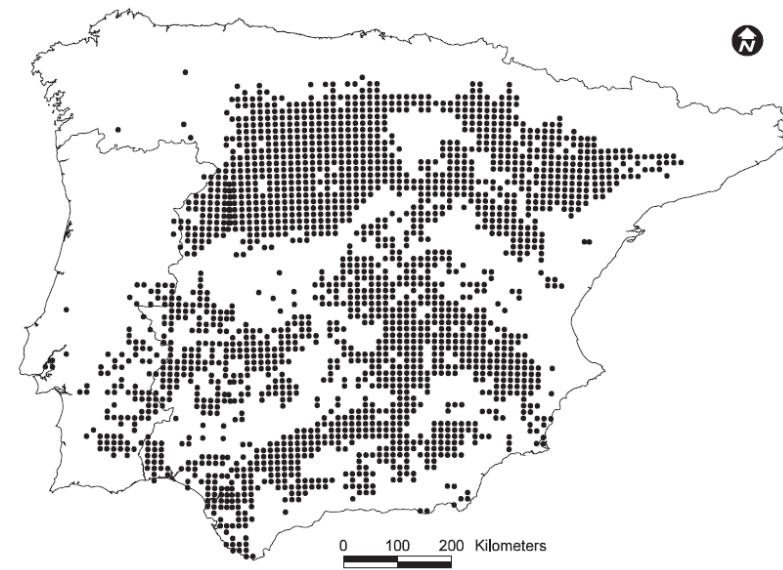
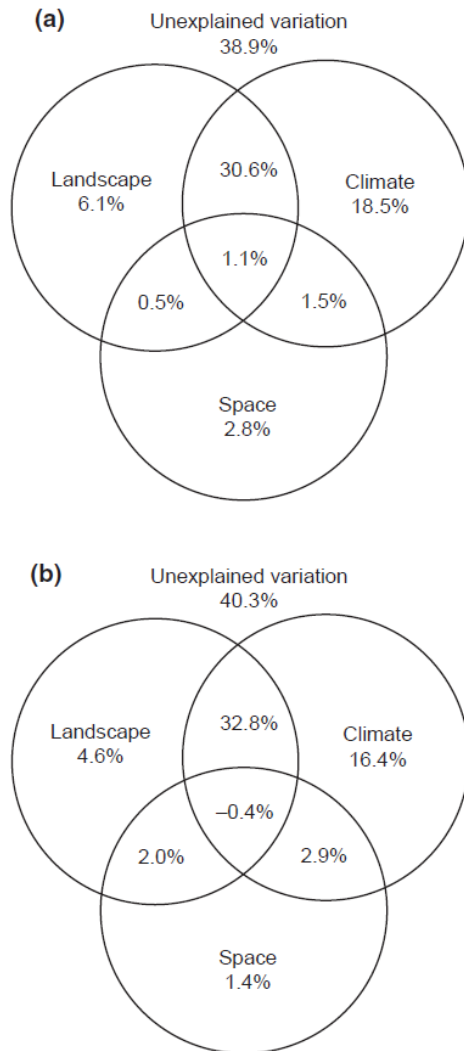
Does local habitat fragmentation affect large-scale distributions? The case of a specialist grassland bird

Luis Reino^{1,2,3*}, Pedro Beja³, Miguel B. Araújo^{1,4,5}, Stéphane Dray⁶ and Pedro Segurado²



Table 1 Summary statistics (mean ± standard deviation) of predictor variables considered in the analyses of factors affecting the presence or absence of the calandra lark in 10 × 10 km squares across the Iberian Peninsula

Variables (Abbreviation; Units)	Presence	Absence
Total annual precipitation (PREC; mm)	504.0 ± 102.4	730.6 ± 276.3
Mean annual air temperature (TANN; °C)	12.9 ± 23.4	12.6 ± 30.3
Annual temperature range (TRAN; °C)	28.6 ± 26.1	24.9 ± 40.6
Mean slope (SLOP;%)	3.6 ± 2.4	8.5 ± 5.4
Habitat area (AREA; km ²)		
Total	29.9 ± 25.2	6.9 ± 11.3
Effective	11.9 ± 21.6	1.5 ± 5.4
Number of patches (NUMP; N)		
Total	7.5 ± 5.8	7.5 ± 7.1
Effective	6.2 ± 4.3	5.2 ± 5.03
Edge density (EDEN; km ha ⁻¹)		
Total	0.59 ± 0.31	0.25 ± 0.28
Effective	0.72 ± 0.34	0.40 ± 0.37



kalandra zpěvná (*Melanocorypha calandra*)

Figure 3 Variation partitioning Venn diagrams representing the pure and shared contributions of climatic/topographic, landscape and spatial sets of variables to the explained variation in the distribution of calandra lark in the Iberian Peninsula. Landscape variables were computed considering either the total (a) or the effective habitat (b).



Literatura

- Legendre, P. & Legendre, L. (2012): Numerical ecology. Third Edition. Elsevier, Amsterdam.
- Borcard, D., Gillet, F. & Legendre, P. (2011): Numerical ecology with R. Springer, New York.
- Borcard, D., Legendre, P. & Drapeau, P. (1992): Partialling out the spatial component of ecological variation. *Ecology*, 73: 1045–1055