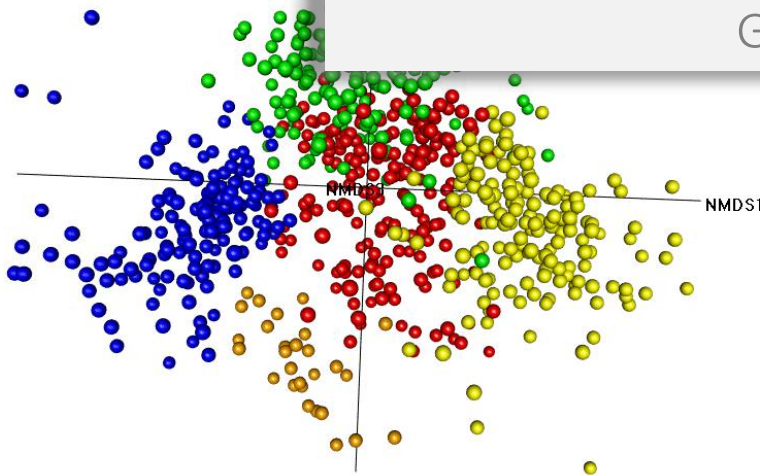


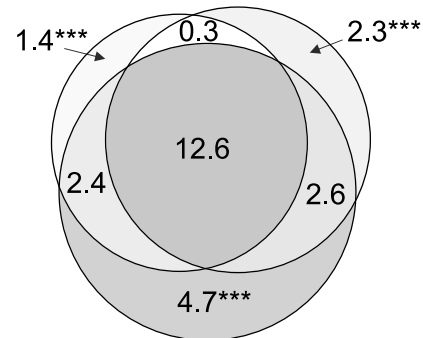
Metody fyzické geografie 3: Biogeografie & ekologie

Jan Divíšek

Geografický ústav & Ústav botaniky a zoologie



climate: 16.6*** land-cover: 17.8***



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



Metody fyzické geografie 3 – 24. 10. 2017

- Teoretická část
 - Příprava dat pro analýzy – EDA, transformace
 - Korelace
 - Regresní analýza – lineární regresní modely a GLM
- Praktická část
 - Import dat do R, sumární statistiky, boxploty, histogramy
 - Transformace dat
 - Korelační analýzy
 - Lineární regrese

Explorativní analýza, transformace a standardizace dat

Data v biogeografii a ekologii

- **Vysvětlovaná proměnná (*Dependent variable(s)*)**

- Distribuce druhů (přítomnost/nepřítomnost), abundance, složení společenstva, vlastnosti druhů atp.
- společenstvo je typicky sledováno na určité ploše (v případě rostlin a některých málo mobilních živočichů) nebo např. inventarizací jedinců (např. ulovených v pastech v případě mobilních živočichů)
- složení živého společenstva je popsáno přítomností jednotlivých druhů daného typu organismů, na jedné ploše (v jedné pasti) se většinou vyskytuje více než jeden druh

- **Vysvětlující proměnná (*Explanatory variable(s)*)**

- Environmentální faktory, vzdálenosti, fylogenetická podobnost atp.
- Prostředí je popisováno jednou nebo více proměnnými, o kterých se předpokládá, že ovlivňují studovaný typ organismů

- **Jednorozměrná data (*univariate data*)**

- pouze jedna proměnná, např. počet druhů

- **Vícerozměrná data (*multivariate data*)**

- matice dat (data matrix), např. lokality × druhy

Typy proměnných

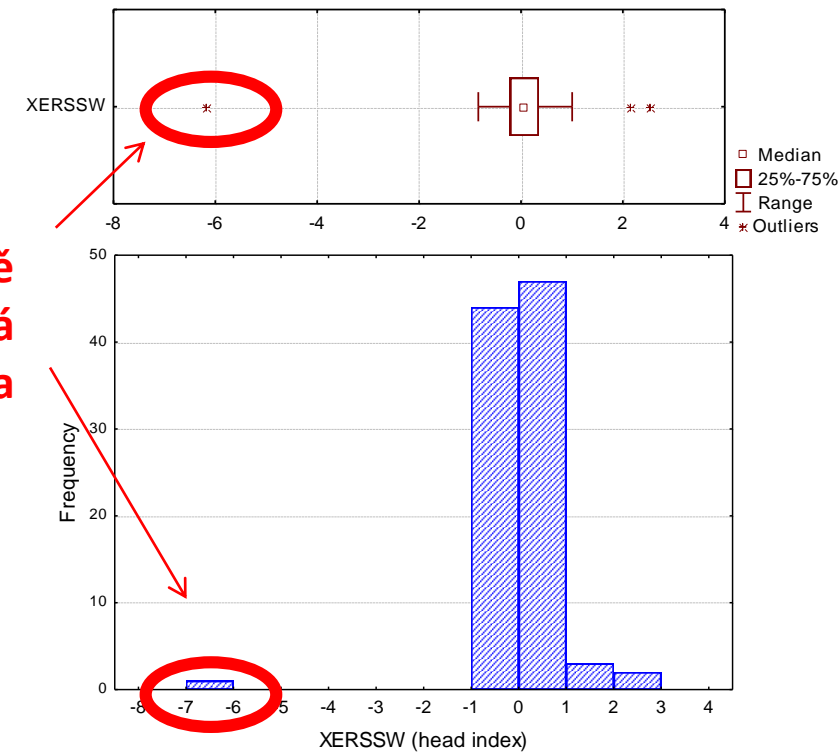
| Typ proměnné | Příklady |
|--|-------------------------------|
| binární (dvoustavový, presence-absence) | přítomnost nebo absence druhu |
| mnohostavový | |
| <ul style="list-style-type: none"> ┆ neseřazený | geologický substrát |
| <ul style="list-style-type: none"> ┆ seřazený | |
| <ul style="list-style-type: none"> ┆┆ semikvantitativní (ordinální) | stupnice pokryvností druhů |
| <ul style="list-style-type: none"> ┆┆┆ kvantitativní (měření) | |
| <ul style="list-style-type: none"> ┆┆┆┆ diskontinuální (počty, diskrétní) | počet jedinců |
| <ul style="list-style-type: none"> ┆┆┆┆┆ kontinuální | teplota, hloubka půdy |

Legendre & Legendre (1998)

Explorační analýza dat (exploratory data analysis, EDA)

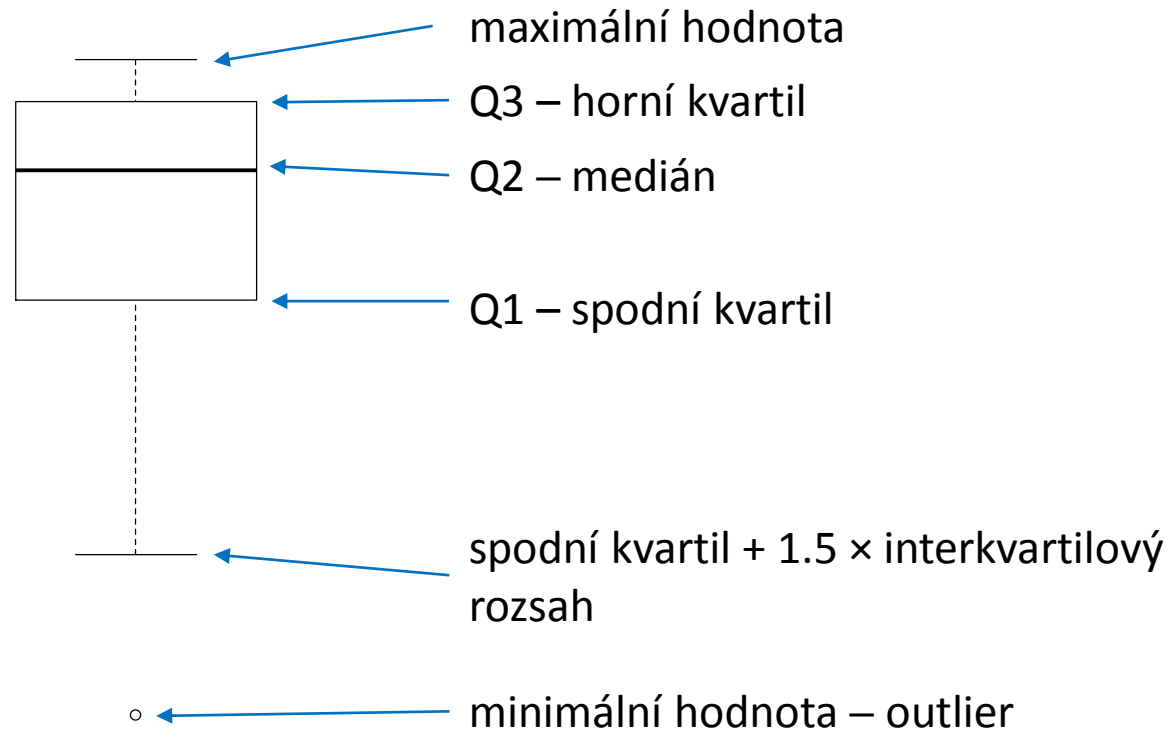
- průzkum dat – kontrola a čištění
 - chyby (*errors*)
 - někdy se chovají jako odlehlé body, je třeba zkontrolovat původní záznam a případně data z analýzy odstranit
 - chybějící data (*missing data, NA*)
 - možnosti jejich nahrazení (interpolace, model)
 - vyloučení proměnné nebo vzorku který má hodně chybějících hodnot
 - odlehlé body (*outliers*)
 - jejich detekce (*outlier analysis*)
- hledání hypotéz, které stojí za to testovat
- **grafická EDA** slouží k
 - odhalení odlehlých bodů (*outlier analysis*)
 - distribuce dat (normalita) a nutnost transformace

**potenciálně
chybná
hodnota**



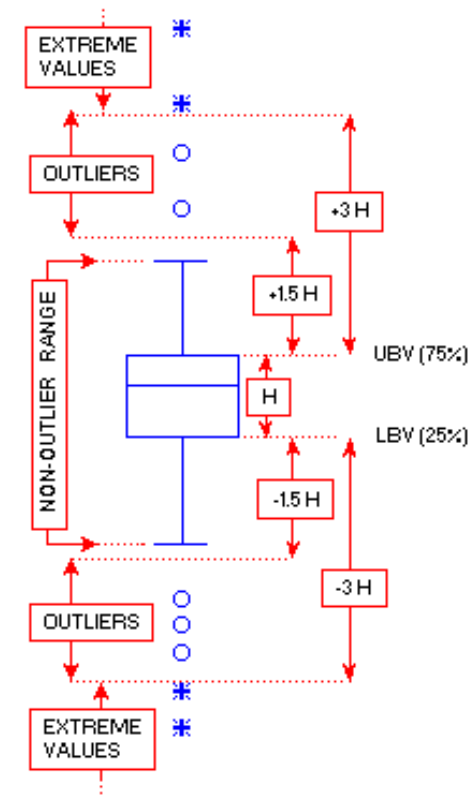
Krabicové grafy (boxplots)

Klasický boxplot (střední hodnota = medián)

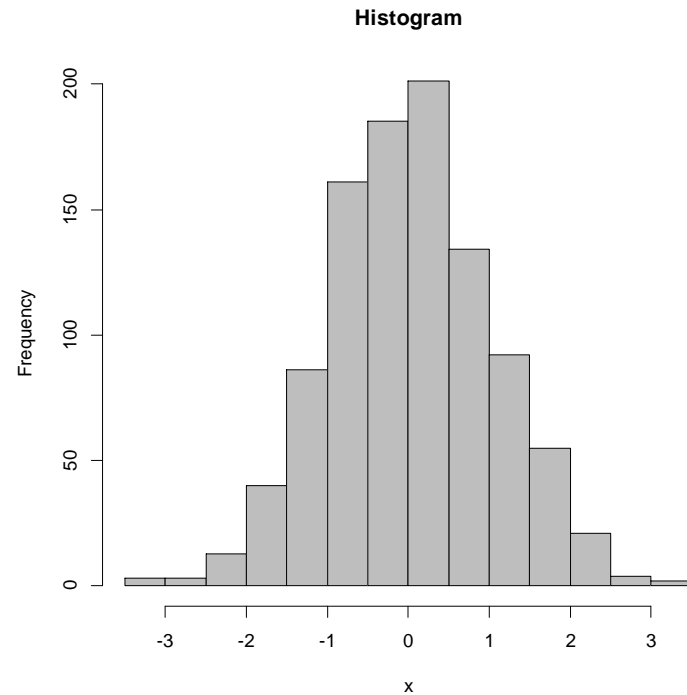


`boxplot()`

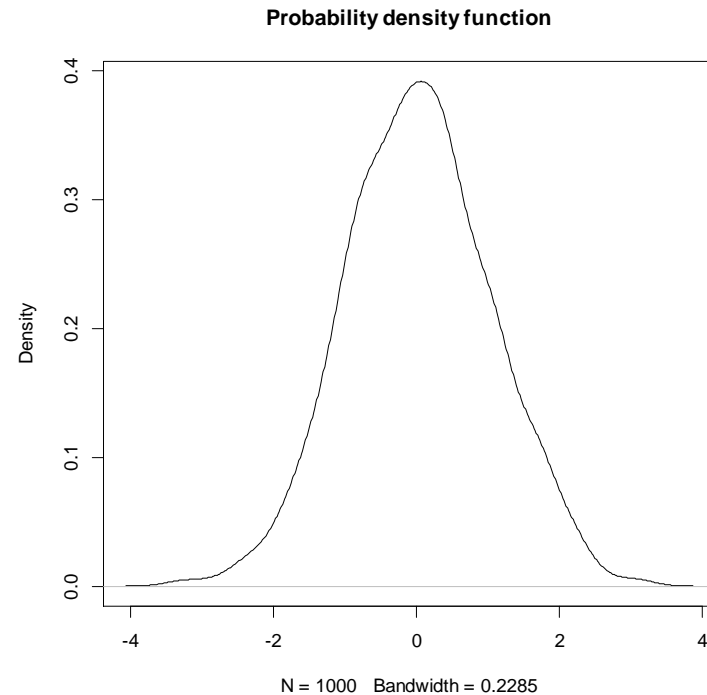
Definice odlehlých bodů a extrémů (STATISTICA)



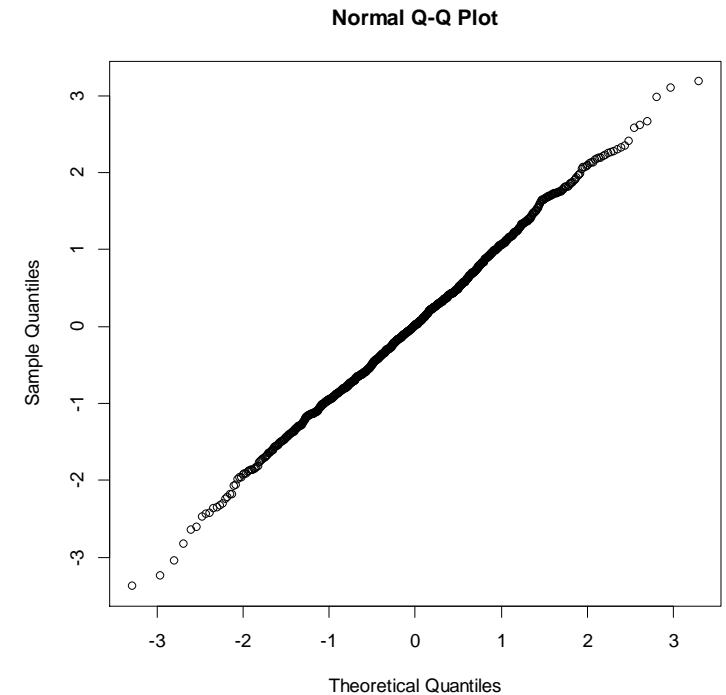
Histograms, PDF plots & Q-Q plots



`hist()`



`density()`



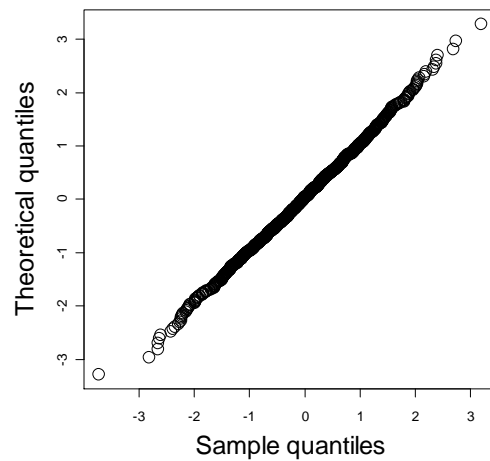
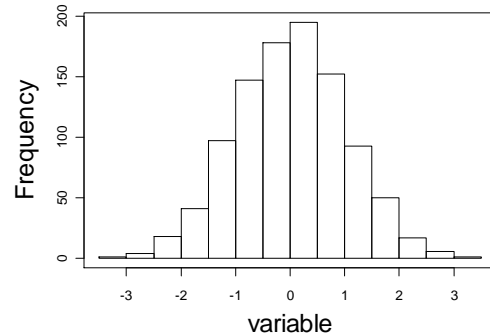
`qqnorm()`

Testování normality dat:

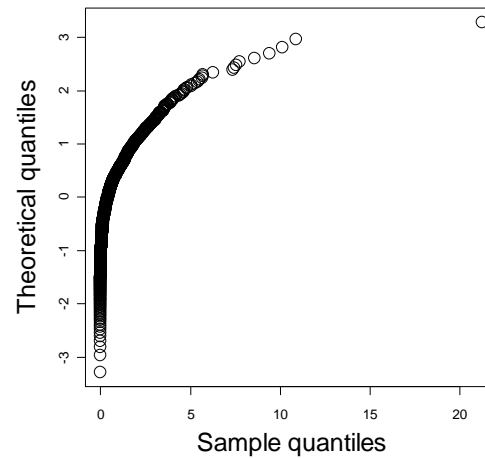
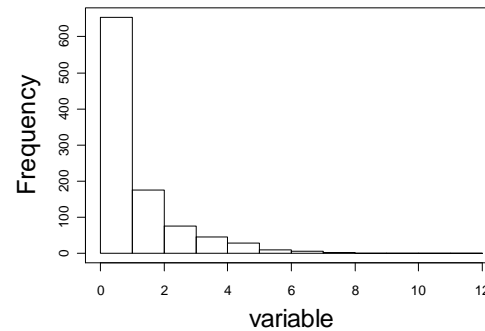
- Shapiro-Wilkův test: `shapiro.test()`
- Kolmogorovův-Smirnovův test: `ks.test()`

Mají data normální rozložení?

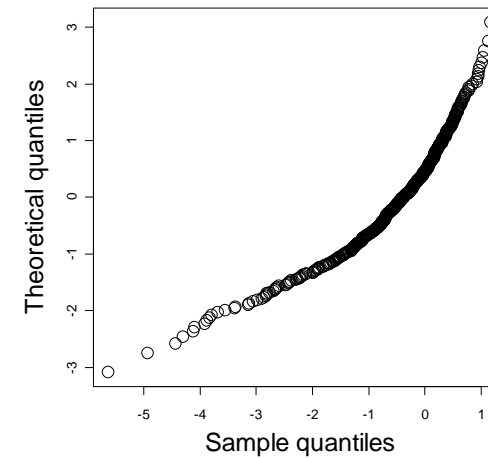
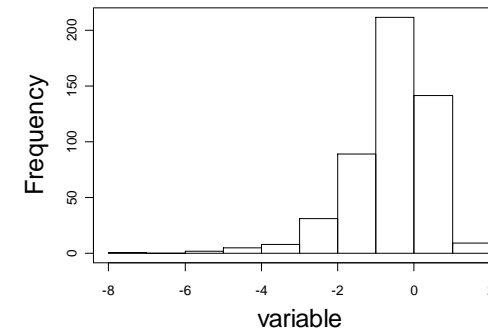
normální rozdělení
(*symetrical*)



pozitivně (doprava) sešikmené
(*right skewed*)



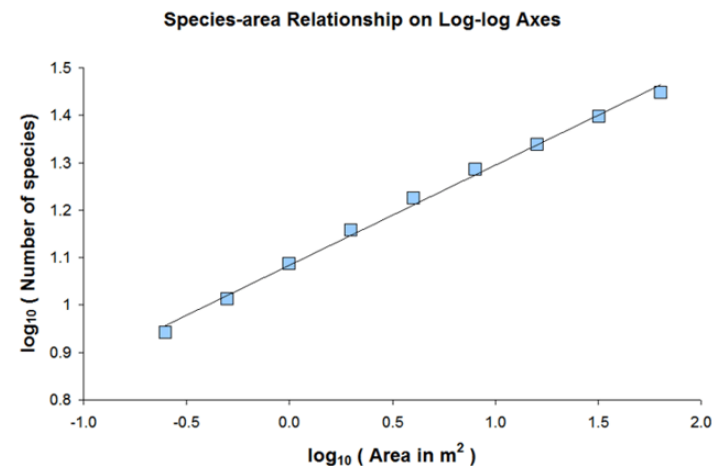
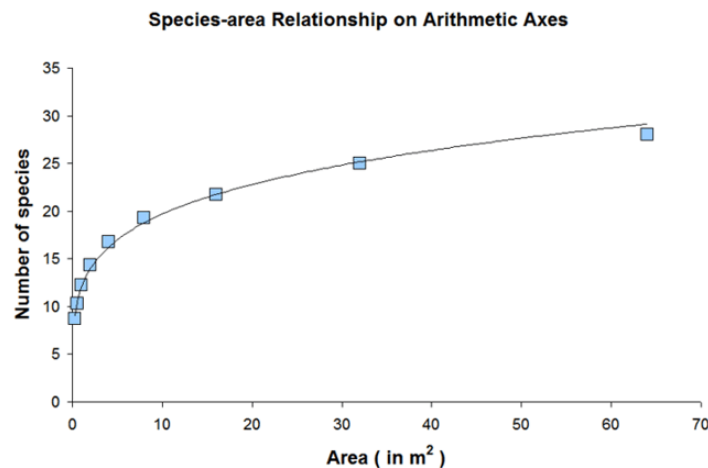
negativně (doleva) sešikmené
(*left skewed*)



ekologická data jsou často zešikmená pozitivně (doprava), protože jsou omezená nulou na začátku

Transformace dat

- mění relativní vzdálenosti mezi jednotlivými hodnotami a tím i tvar jejich distribuce
- **Proč data transformovat?**
 - parametrické testy jsou založené na předpokladu, že data mají nějaké určité (často normální) rozdělení
 - protože lineární vztahy se dají popsat přímkou a lépe se interpretují než vztahy nelineární
 - škála měření je arbitrární a nemusí odpovídat ekologickému významu proměnné (používáme desítkovou soustavu)



https://en.wikipedia.org/wiki/Species-area_curve

Transformace dat

- Na co si dát při transformaci pozor?
 - aby transformace rozložení dat ještě nezhoršila a nevytvořila nové odlehlé body
 - abychom při komentování výsledků používali netransformované hodnoty proměnných
- Typy transformace
 - lineární
 - přičtení konstanty nebo vynásobení konstantou
 - nemění výsledky statistického testování nulových hypotéz
 - např. převod teploty měřené ve stupních Celsia na stupně Fahrenheita
 - nelineární
 - log transformace, odmocninová transformace atd.
 - může změnit výsledky statistického testování

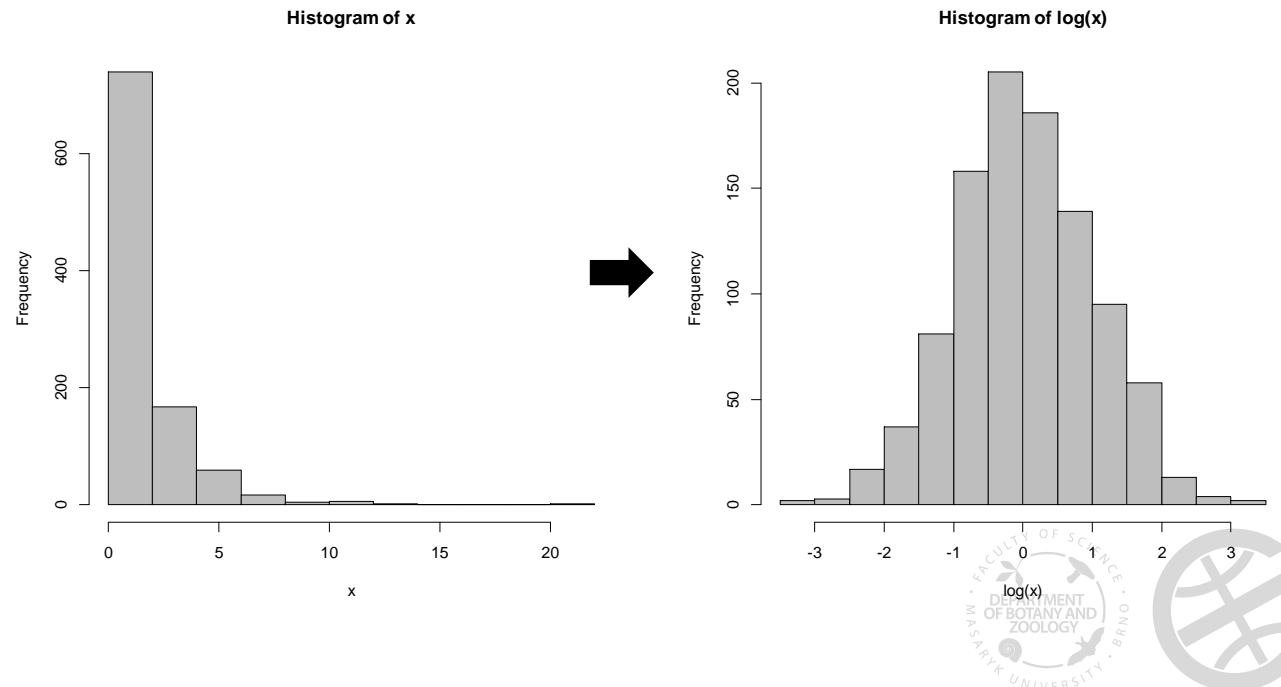
Typy transformací

• Logaritmická transformace (*log transformation*)

- pro data s výrazně pozitivně (doprava) šikmou distribucí (*right skewed*), u kterých existuje vztah mezi směrodatnou odchylkou a průměrem (lognormální rozložení)

$$Y' = \log(Y) \quad \text{případně} \quad Y' = \log(aY + c)$$

- na základě logaritmu nezáleží (10, 2, e)
- konstanta $a = 1$; pokud je Y z intervalu $\langle 0;1 \rangle$, potom $a > 1$
- konstanta c se přidává, pokud proměnná Y obsahuje nuly
- c může být např. 1, nebo arbitrárně zvolené malé číslo (0,001)
- na konstantě c může záležet výsledek analýz (ANOVA), a proto je dobré vybírat takové číslo, aby transformovaná proměnná byla co nejvíce symetrická



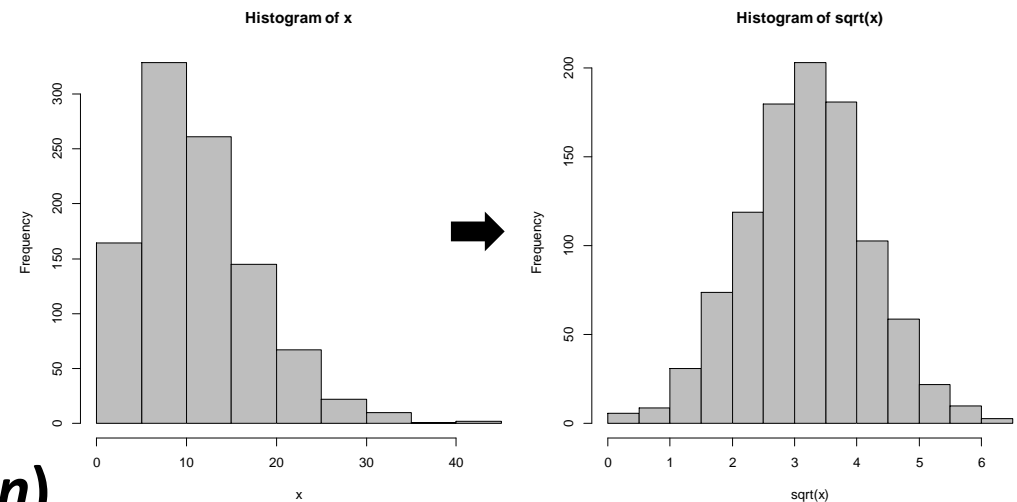
Typy transformací

• Odmocninová transformace (*square-root transformation*)

- vhodná pro mírně doprava zešikmená data (*right skewed*), např. počty druhů (*Poisson distribution*)
- třetí a vyšší odmocnina je účinnější na více zešikmená data (čtvrtá odmocnina se používá pro abundance druhů s mnoha nulami a několika vysokými hodnotami)

$$Y' = \sqrt{Y} \quad \text{případně} \quad Y' = \sqrt{Y + c}$$

- konstanta c se přičítá, pokud soubor obsahuje nuly
- c může být např. 0,5, nebo 3/8 (0,325)



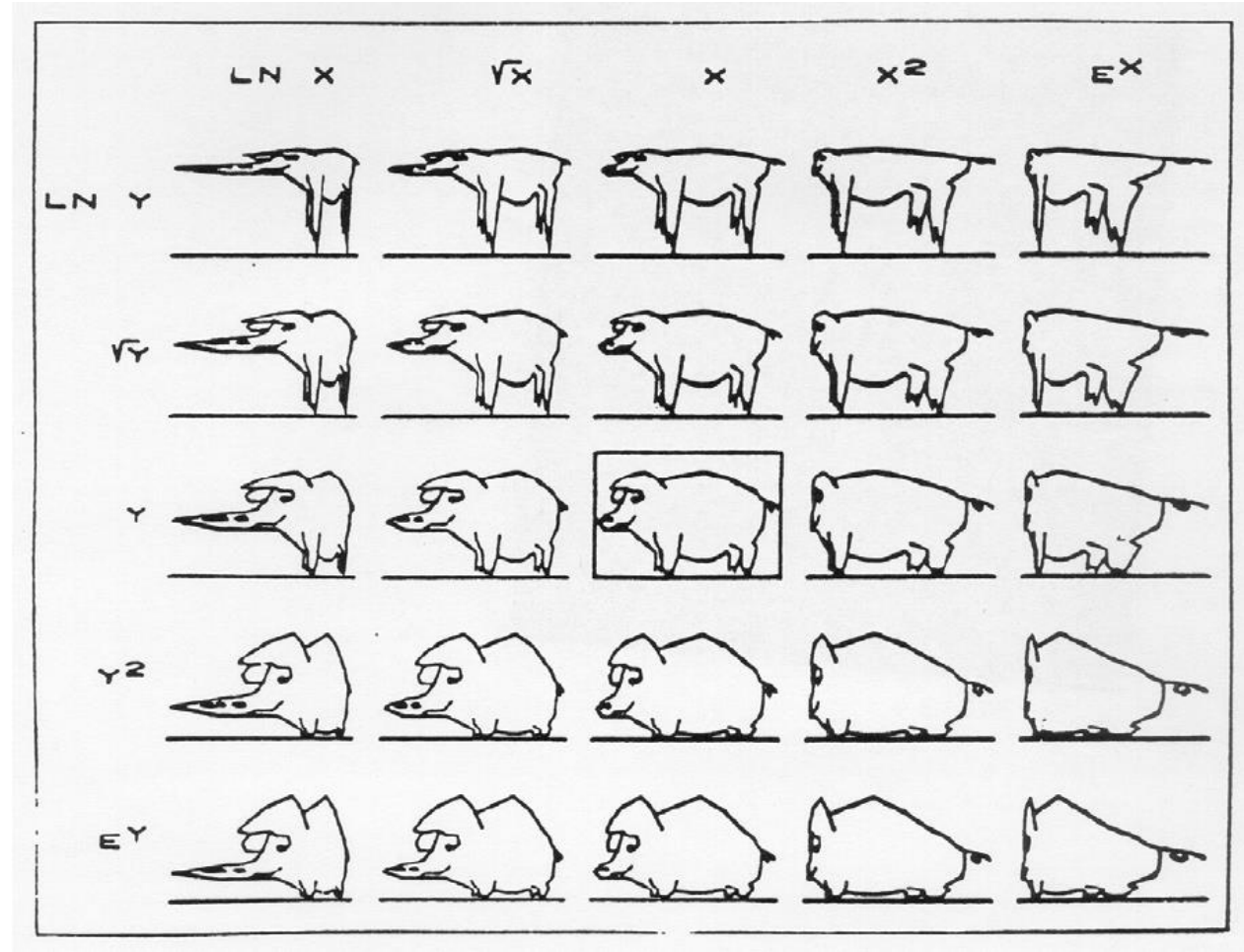
• Mocninná transformace (*power transformation*)

- vhodná pro data negativně (doleva) sešikmená (*left skewed*)

$$Y' = Y^p$$

- pokud $p < 1$ - **odmocninová transformace** ($p = 0,5$ – druhá odmocnina, $p = 0,25$ – čtvrtá odmocnina atd.)

Transformace



Münch. Med. Wschr. 124, 1982

Další transformace

- **Transformace pomocí arcsin (*angular transformation*)**

- vhodná pro procentické hodnoty (a obecně podíly)

$$Y' = \sin Y \quad \text{případně} \quad Y' = \sin \sqrt{Y}$$

- použitelná pro hodnoty v intervalu $\langle -1; 1 \rangle$
- transformované hodnoty jsou v radiánech

- **Reciproká transformace (*reciprocal transformation*)**

- vhodná pro poměry (například výška/hmotnost, počet dětí v populaci na počet žen atd.)

$$Y' = 1/Y$$

- **Box-Cox transformace (zobecněná mocninná transformace)**

- zobecněná parametrická transformace
- iterativní hledání parametru λ (lambda), pro které je rozdělení transformované proměnné nejbližší normálnímu rozdělení
- používá se v případě, že nemáme *a priori* představu, jakou transformaci použít



Standardizace dat

- vyrovnává rozdíly v relativním významu (váze) jednotlivých ekologických proměnných (měřené na různých škálách), druhů nebo vzorků
- mění data pomocí statistiky, která je spočtená na datech samotných, např. průměr, součet, rozsah aj. (*data dependent*)
- ve své podstatě je to další typ transformace

Standardizace dat

- **Centrování (*centring*)**

- výsledná proměnná má průměr roven nule

$$Y'_i = Y_i - \text{průměr (Y)}$$

- **Standardizace v úzkém slova smyslu**

- výsledná proměnná má průměr roven nule a směrodatnou odchylku rovnu jedné
- „synchronizuje“ proměnné měřené v různých jednotkách a na různých stupnicích

$$Y'_i = (Y_i - \text{průměr (Y)}) / \text{směrodatná odchylka (Y)}$$

- **Změna rozsahu hodnot (*ranging*)**

- výsledná proměnná je v rozsahu [0, 1]

$$Y'_i = Y_i / Y_{max} \quad \text{nebo} \quad Y'_i = (Y_i + \text{abs}(Y_{min})) / (Y_{max} + \text{abs}(Y_{min}))$$



Kódování dat

- *Dummy variables*

- metoda, jak převést **kvalitativní** (kategoriální) proměnnou na **kvantitativní** (binární) proměnné použitelné v analýzách
- pokud má kategoriální proměnná n stavů (hodnot), pro její vyjádření stačí $n-1$ dummy proměnných (jedna z proměnných je vždy lineárně závislá na ostatních)
- `dummy { dummies }`

| hodnoty | dummy proměnné | | | |
|----------|----------------|------|------|-------|
| | KAMB | LITO | RANK | FLUVI |
| kambizem | 1 | 0 | 0 | 0 |
| litozem | 0 | 1 | 0 | 0 |
| ranker | 0 | 0 | 1 | 0 |
| fluvizem | 0 | 0 | 0 | 1 |

Kódování dat

- např. nahrazení kódů u alfa-numerických stupnic, např. Braun-Blanquetovy stupnice dominance-abundance

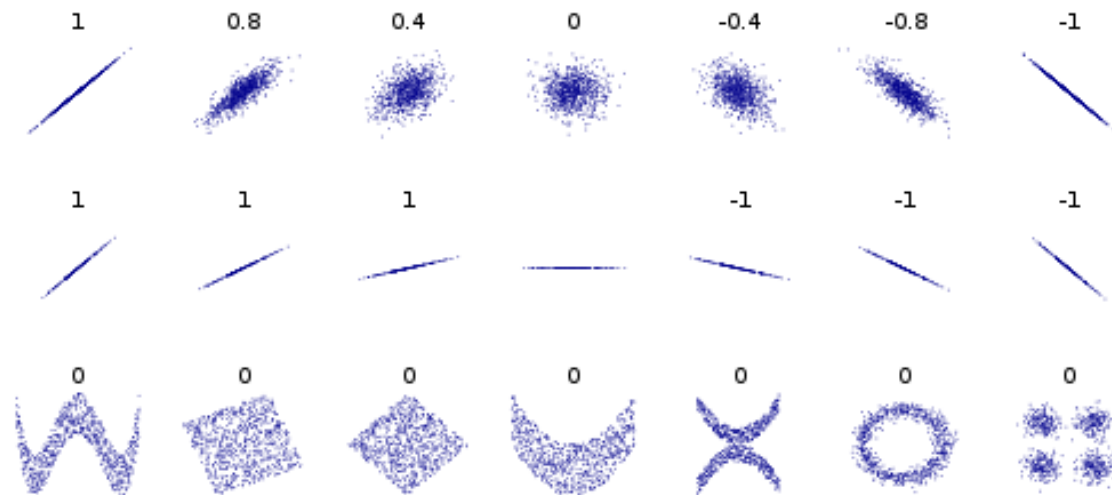
| | | | | | | | |
|-----------------------------|----------|----------|----------|----------|----------|----------|----------|
| Braun-Blanquetova stupnice: | r | + | 1 | 2 | 3 | 4 | 5 |
| ordinální hodnoty: | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| střední hodnoty procent: | 1 | 2 | 3 | 13 | 38 | 63 | 88 |

Korelační a regresní analýza

Korelační analýza

`cor()`
`cor.test()`

- Korelace = vzájemný vztah mezi dvěma procesy nebo veličinami
- Ve statistice popisuje vzájemný lineární vztah mezi veličinami x a y
- Míru korelace vyjadřuje **korelační koeficient**, který může nabývat hodnot od -1 až po $+1$.
 - Pearsonův korelační koeficient (r)
 - Spearmanův koeficient pořadové korelace (ρ nebo r_s)



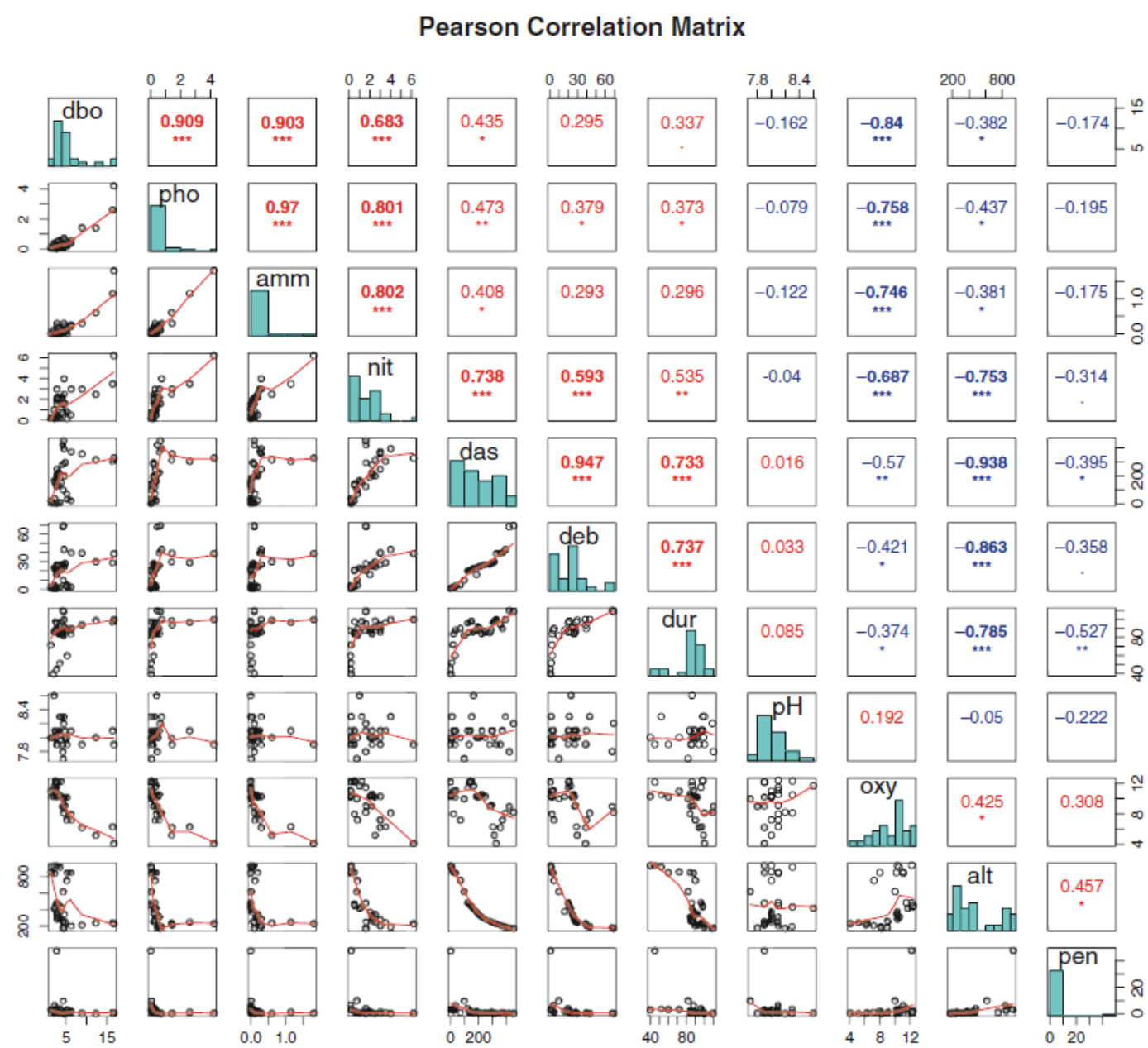


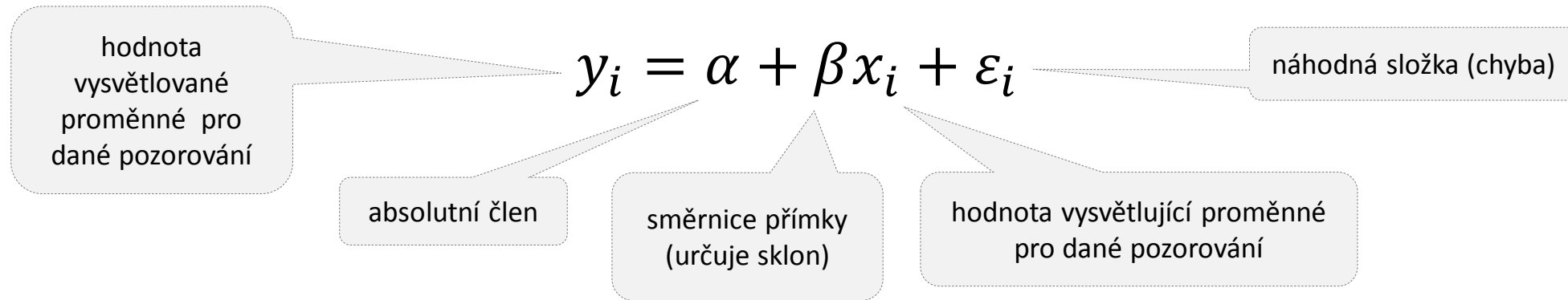
Fig. 3.3 Multipanel display of pairwise relationships between environmental variables with Pearson's r correlations

Borcard et al. (2011)

Lineární regrese

lm()

- Lineární regresní model popisuje vztah mezi jednou závislou proměnnou a jednou nebo více vysvětlujícími proměnnými



$$\epsilon_i \sim N(0, \sigma^2), \text{cor}(\epsilon_i, \epsilon_{i'}) = 0 \text{ pro } i \neq i'$$

chyby mají normální rozdělení (N) s nulovou střední hodnotou (0) a rozptylem (σ^2) který je stejný pro všechna pozorování

vzájemná korelace chyb je nulová

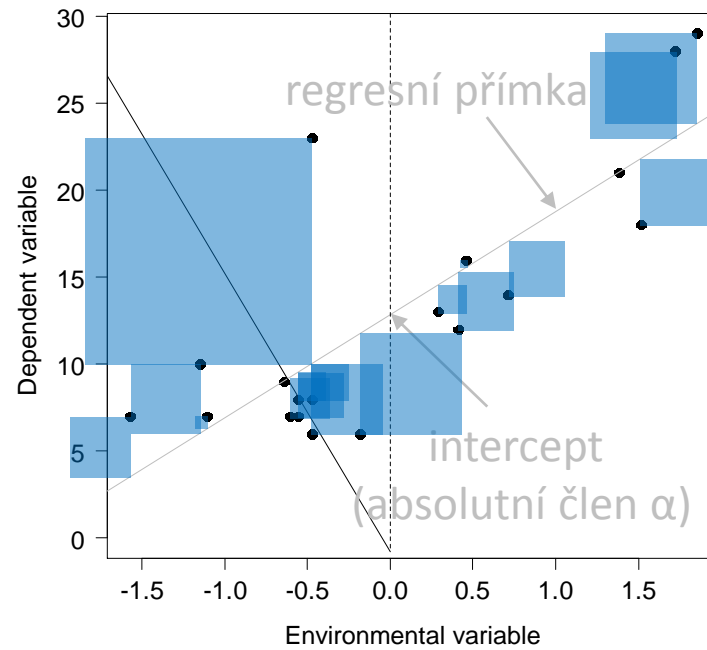
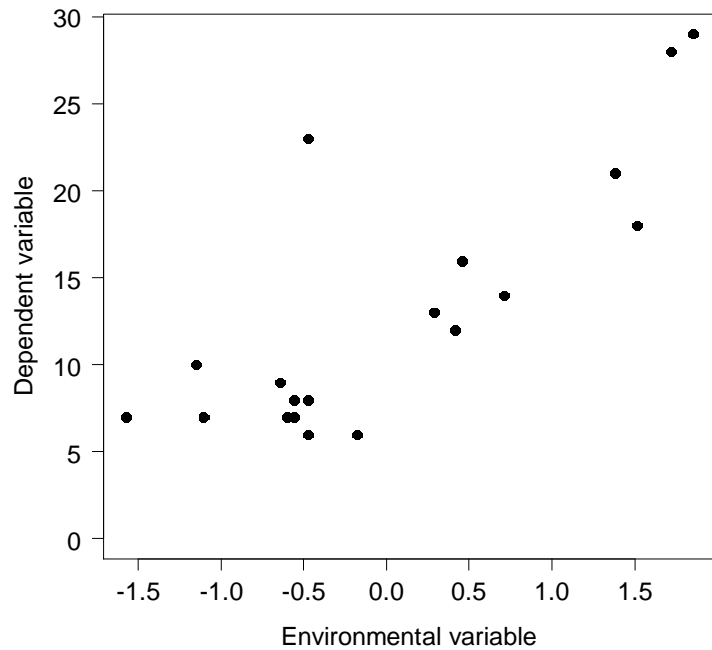
Lineární regrese

- Lineární regresní model popisuje vztah mezi jednou závislou proměnnou a jednou nebo více vysvětlujícími proměnnými

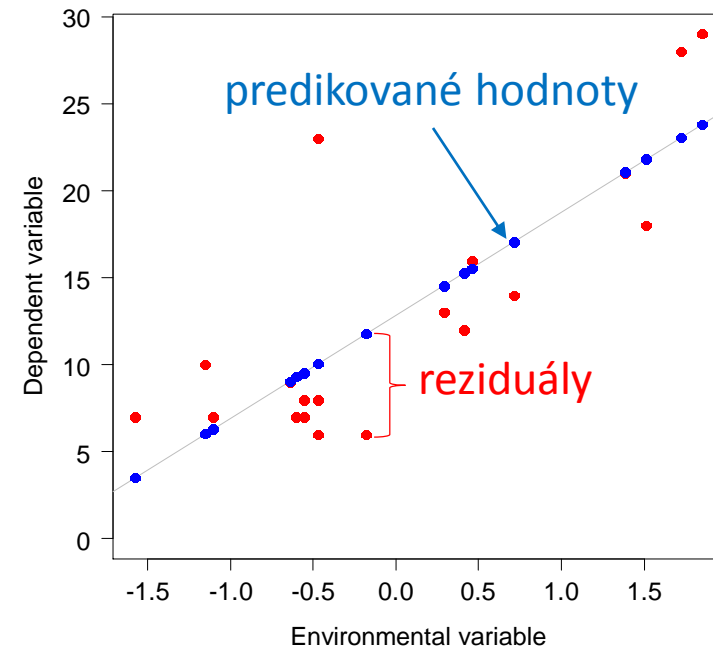
$$y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon$$

kde $\varepsilon \sim N(0, \sigma^2)$, nezávisle pro různá měření

Lineární regrese



`lm()`

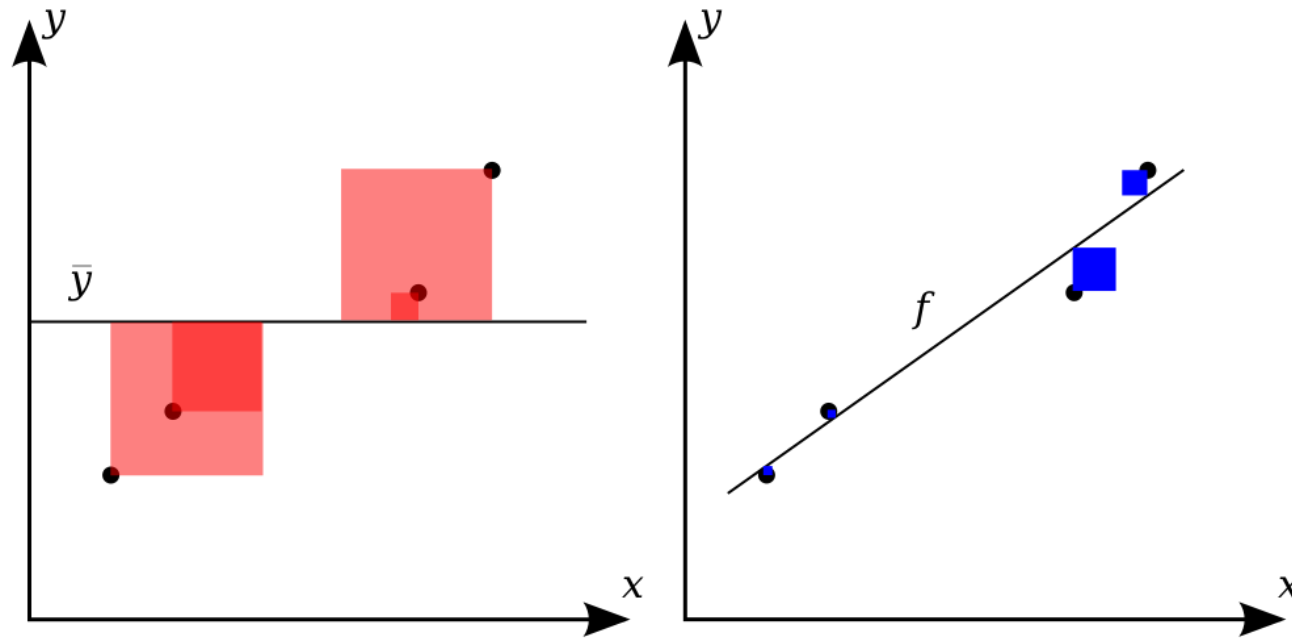


`predict()`

`resid()`

Vysvětlená variabilita v regresi

$$TSS = \sum_{i=1}^n (y_i - \bar{y})^2$$



$$RSS = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

$$R^2 = 1 - \frac{RSS}{TSS}$$

Lineární regresní model v R

```
m <- lm(Y ~ X)
```

```
anova(m)
```

Analysis of Variance Table

stupně volnosti
(Degrees of freedom)

Response: Species

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|-----------|----|--------|---------|---------|---------------|
| Altitude | 1 | 503.89 | 503.89 | 17.488 | 0.0005604 *** |
| Residuals | 18 | 518.66 | 28.81 | | |

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

průměrné sumy čtverců (Sum Sq/Df)

hodnoty F-testového kritéria
(podíl Mean Sq pro danou
proměnnou a Mean Sq pro
rezidua)

statistická významnost

sumy čtverců (Sum of Squares)

pro vysvětlující proměnné (model sum of squares): $MSS = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$

pro reziduály (residual sum of squares): $RSS = \sum_{i=1}^n (y_i - \hat{y}_i)^2$

$TSS = MSS + RSS$

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 |

Lineární regresní model v R

`summary(m)`

Call:
lm(formula = Species ~ ., data = dat)

Residuals:

| Min | 1Q | Median | 3Q | Max |
|---------|---------|--------|--------|--------|
| -7.4215 | -1.9238 | 0.6839 | 2.3229 | 6.6599 |

střední chyba

Coefficients:

odhady koeficientů

| | Estimate | Std. Error | t value | Pr(> t) |
|-------------|----------|------------|---------|--------------|
| (Intercept) | 12.8500 | 0.8962 | 14.338 | 9.24e-10 *** |
| Altitude | -2.0206 | 1.4600 | -1.384 | 0.1880 |
| Slope | 0.1462 | 1.0439 | 0.140 | 0.8906 |
| pH | 4.0380 | 1.5358 | 2.629 | 0.0198 * |
| Moisture | 1.6358 | 1.0076 | 1.623 | 0.1268 |
| E3_cover | 1.6526 | 1.1758 | 1.405 | 0.1817 |

hodnoty t-testu nulové hypotézy (H0) o tom, že skutečná hodnota daného koeficientu je nulová

statistická významnost

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

variabilita v Y vysvětlená modelem (R^2)

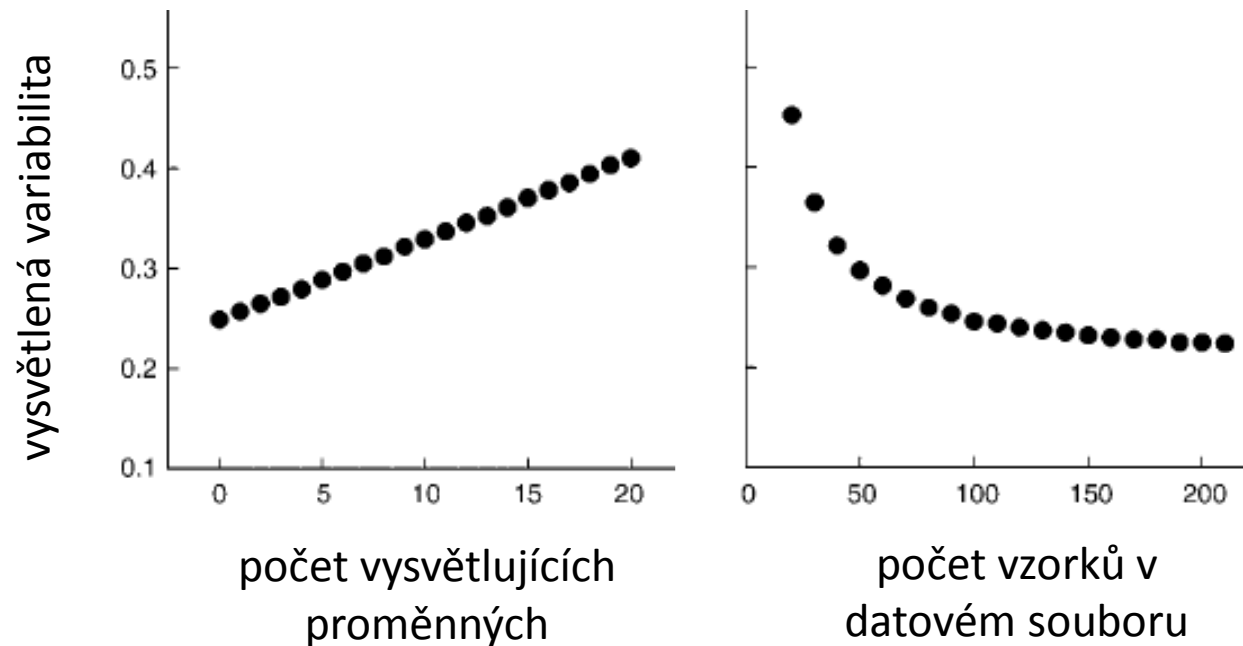
Residual standard error: 4.008 on 14 degrees of freedom
Multiple R-squared: 0.78, Adjusted R-squared: 0.7015
F-statistic: 9.93 on 5 and 14 DF, p-value: 0.0003219

adjustovaný R^2



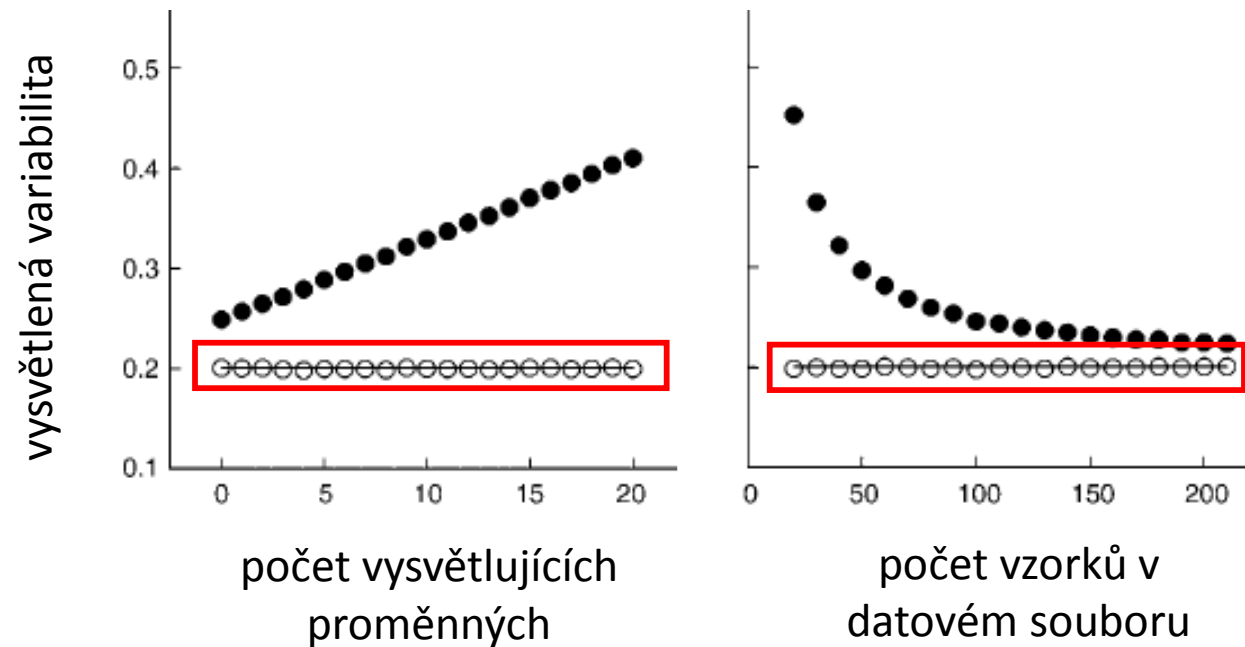
Vysvětlená variabilita (R^2)

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



Vysvětlená variabilita (R^2) a **adjustovaný R^2**

- **adjustovaný R^2 se nemění s počtem vysvětlujících proměnných a počtem vzorků v souboru**



Výpočet adjustovaného R^2

`RsquareAdj` {vegan}

- pomocí Ezekielovy formule

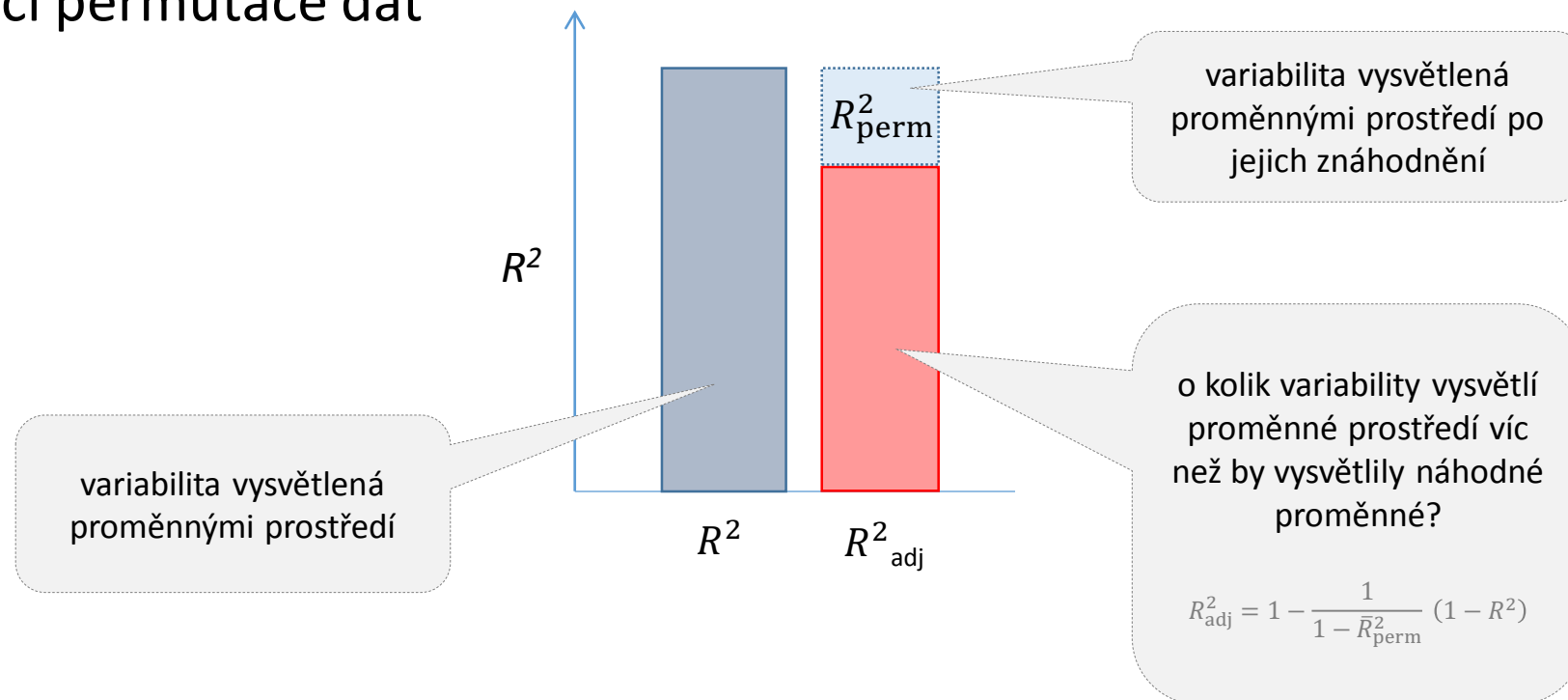
$$R^2_{(Y|X)adj} = 1 - \frac{n-1}{n-p-1} (1 - R^2_{Y|X})$$

n ... počet vzorků

p ... počet vysvětlujících proměnných

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

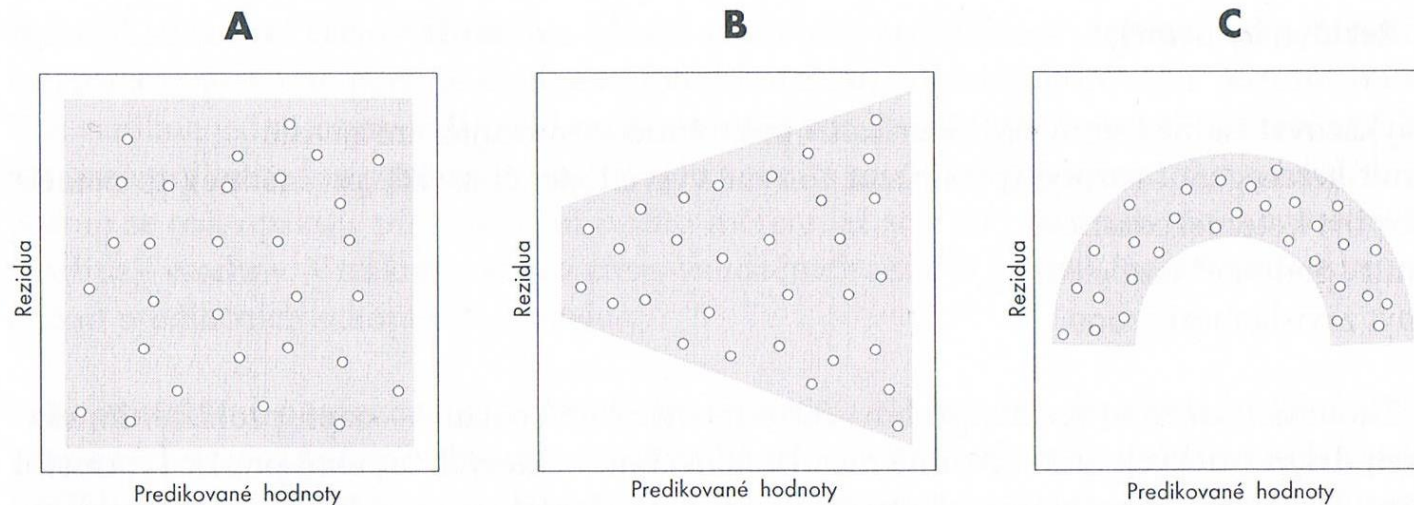
- pomocí permutace dat



Rezidua lineárního regresního modelu

`resid()`

- Rezidua by neměla:
 - vykazovat trendy vůči kterékoliv proměnné, vysvětlující ani závislé
 - mít heterogenní rozptyl (přes různé úrovně vysvětlující či závislé proměnné), tj. neměla by být heteroskedastická
 - mít „podivné“ rozdělení (předpokládá se normální)
 - být závislá mezi sebou (autokorelovaná)



Obr. 4-1 Závislost reziduí na predikovaných hodnotách. **A.** Homogenní rozptyl. **B.** Rozptyl rostoucí se střední hodnotou. **C.** Zakřivený trend v reziduích. Pro názornost je trend zvýrazněn šedou plochou.

Pekár & Brabec (2009)

Zobecněné lineární modely (GLM)

`glm()`

- Umožňují modelovat proměnné, které nesplňují předpoklady lineárního modelu
 - „nenormální“ rozložení dat (Lognormální, Poissonovo, Binomické atp.)
 - rozptyl se mění s průměrem
 - ...

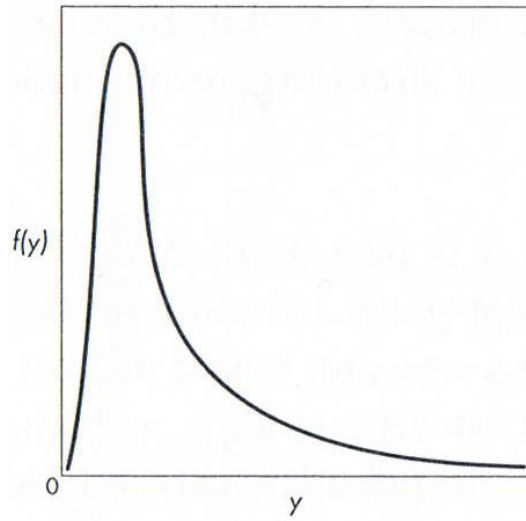
$$\eta_i = \alpha + \sum_{j=1}^p \beta_j x_{ji} \quad y_i = \hat{y}_i + \varepsilon_i, \text{ kde } g(\hat{y}_i) = \eta_i$$

kanonická link funkce

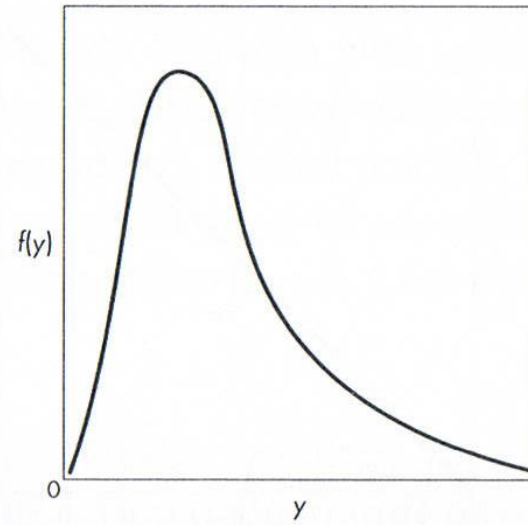
- Parametry GLM
 - Transformační funkce (link) – volí se podle typu rozložení dat
 - Lineární prediktor
 - Náhodná složka

Typy rozdělení

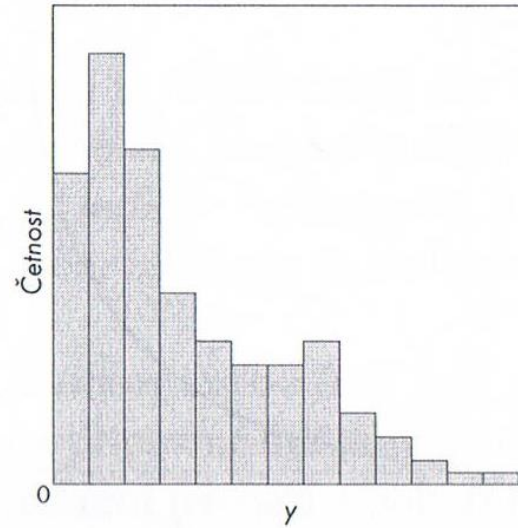
Lognormální rozdělení



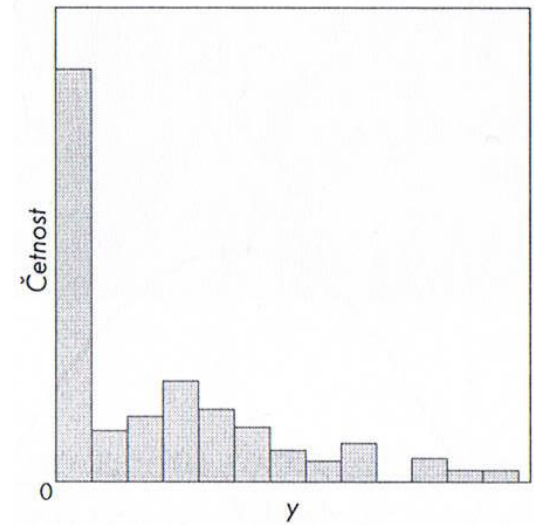
Gamma rozdělení



Poissonovo rozdělení

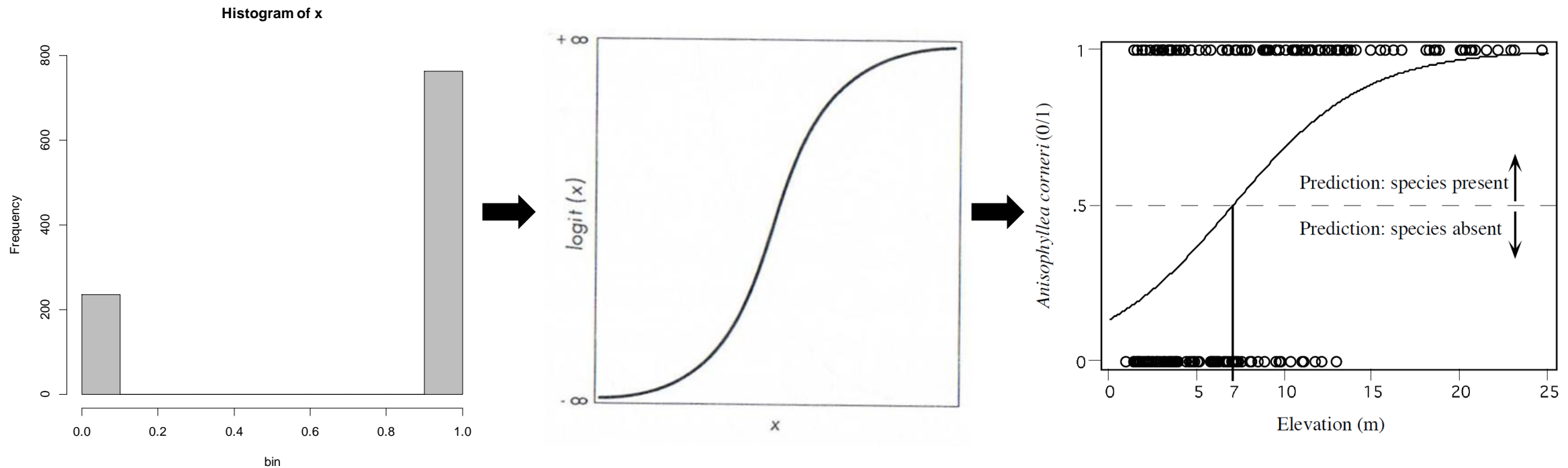


Negativně binomické rozdělení



Binární proměnné

- Např. přítomnost (1) × nepřítomnost (0) druhu
- Binární proměnná + GLM + logit link → **logistická regrese (*logistic regression*)**



Legendre & Legendre (2012)

Kanonické link funkce

| Rozdělení | Jméno linku | Link funkce | Rozptyl | Vysvětlovaná proměnná | |
|---------------------|-------------|------------------------------------|----------------------------------|-----------------------|---|
| | | | | Hodnoty | Typy údajů |
| Gaussovo (normální) | identity | \hat{y} | 1 | jakékoliv reálné | s ohledem na ostatní možnosti skoro žádné |
| Gamma | inverse | $\frac{1}{\hat{y}}$ | \hat{y}^2 | kladné reálné | velikosti, hmotnosti, jejich podíly |
| Poissonovo | log | $\log(\hat{y})$ | \hat{y} | celé nezáporné | počty případů |
| Binomické | logit | $\log \frac{\hat{y}}{1 - \hat{y}}$ | $\frac{\hat{y}(1 - \hat{y})}{n}$ | podíly z počtů | pravděpodobnosti jevů či výsledků |

Pekár & Brabec (2009); Šmilauer (2007)

GLM v prostředí R

```
m <- glm(Y ~ X1 + X2 + ... + X5, family = poisson)
```

```
anova(m, test="Chisq")
```

```
Analysis of Deviance Table
```

```
Model: poisson, link: log
```

```
Response: Species
```

```
Terms added sequentially (first to last)
```

stupně volnosti
(Degrees of freedom)

| | Df | Deviance | Resid. | Df | Resid. Dev | Pr(>Chi) |
|----------|----|----------|--------|--------|------------|----------|
| NULL | | | 19 | | 71.975 | |
| Altitude | 1 | 41.217 | 18 | 30.758 | 1.363e-10 | *** |
| Slope | 1 | 0.454 | 17 | 30.305 | 0.500606 | |
| pH | 1 | 10.081 | 16 | 20.224 | 0.001498 | ** |
| Moisture | 1 | 5.241 | 15 | 14.983 | 0.022061 | * |
| E3_cover | 1 | 0.720 | 14 | 14.263 | 0.396138 | |

```
---
```

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

analogie součtu čtverců (MSS) v LM;
pro GLM s Gaussovým linkem
totožné se součtem čtverců; jinak se
výpočet liší

analogie reziduálnímu součtu
čtverců (RSS) v LM; pro GLM s
Gaussovým linkem totožné se
součtem čtverců; jinak se výpočet
liší

Výsledek χ^2 testu (pokud
Poissovo rozdělení); pokud
byl vhodnější F-test nutné
specifikovat v argumentu
'test'

GLM v prostředí R

```
m <- glm(Y ~ X1 + X2 + ... + X5, family = poisson)
```

```
summary(m)
```

```
...
```

Coefficients:

odhady koeficientů a jejich chyby; pokud je použita logaritmická link funkce lze převést na jednotky Y pomocí exponenciální funkce `exp()`

| | Estimate | Std. Error | z value | Pr(> z) | |
|-------------|----------|------------|---------|----------|-----|
| (Intercept) | 2.44461 | 0.06888 | 35.490 | <2e-16 | *** |
| Altitude | -0.21654 | 0.11605 | -1.866 | 0.0620 | . |
| Slope | -0.03097 | 0.06996 | -0.443 | 0.6581 | |
| pH | 0.24419 | 0.11070 | 2.206 | 0.0274 | * |
| Moisture | 0.13165 | 0.07494 | 1.757 | 0.0789 | . |
| E3_cover | 0.07156 | 0.08460 | 0.846 | 0.3976 | |

tzv. Waldovy statistiky – jejich předpoklady často nejsou splněny

statistická významnost

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

analogie TSS

```
(Dispersion parameter for poisson family taken to be 1)
```

analogie RSS

```
Null deviance: 71.975 on 19 degrees of freedom
Residual deviance: 14.263 on 14 degrees of freedom
```

Akaikeho informační kritérium

```
AIC: 111.69
```

```
Number of Fisher Scoring iterations: 4
```


Rezidua v GLM

`resid()`

- U modelů s jiným než identickým linkem (Gaussovo rozdělení) je několik typů reziduí
 - Rezidua na transformované škále (např. log při Poissonově rozdělení): `type = "working"`
 - Pearsonova rezidua (obdoba standardizovaných reziduí v LM): `type = "pearson"`
 - Prostá rezidua na původní škále: `type = "response"`

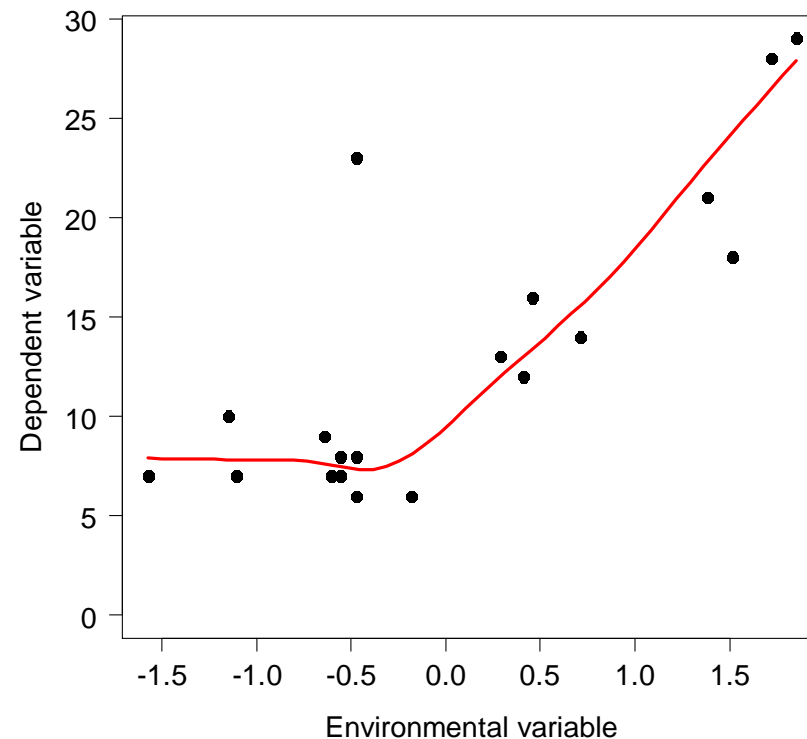


Další vychytávky (nelineární trendy)

- **Lowess and loess smoothing methods**

- Neparametrický odhad trendu pořízený na základě velmi flexibilní lokální regrese
- Fituje křivku na data → dobré pro ukázání vztahu proměnných
- Vhodné jen pro $n < 1000$, pokud více → GAM

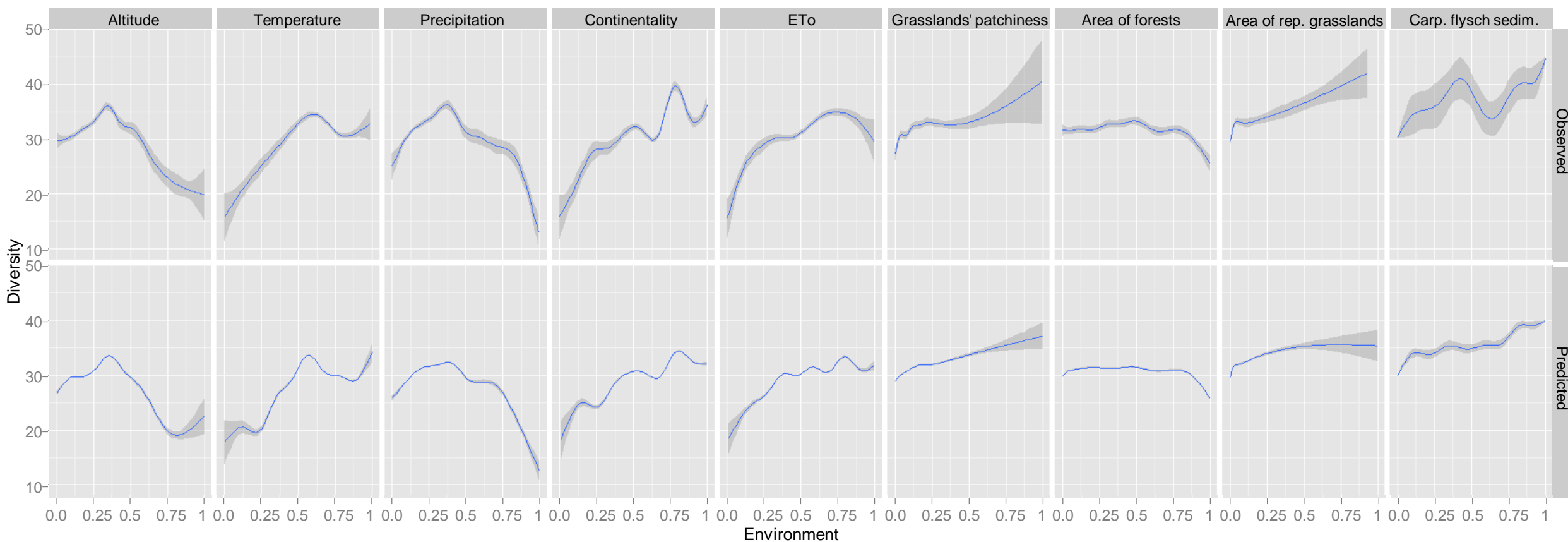
```
loess.smooth()  
lowess()
```



Další vychytávky (nelineární trendy)

- Generalized Additive Models (GAM)

gam()



Literatura

- Pekár, S. & Brabec, M. (2009): Moderní analýza biologických dat. Scientia, Praha.
- 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.