

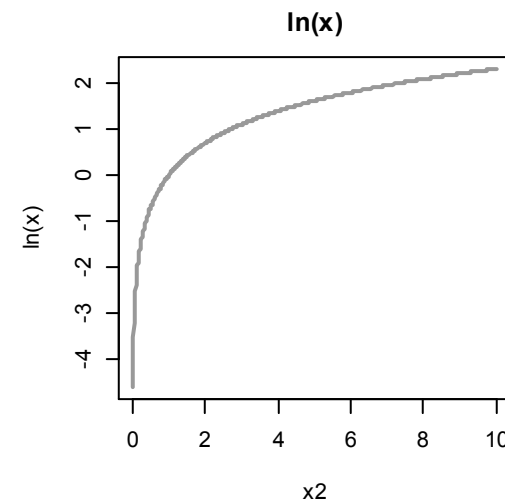
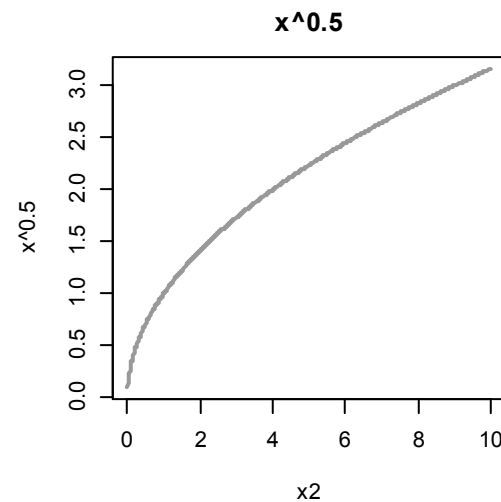
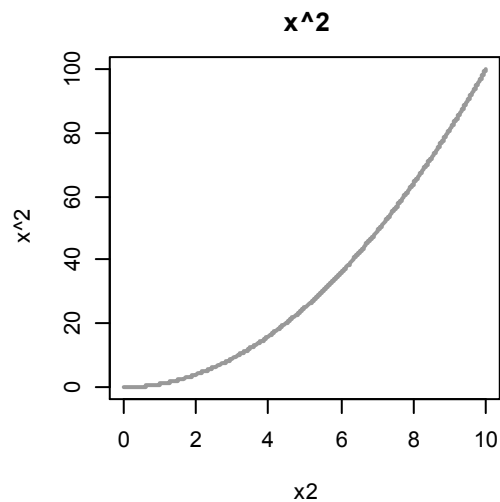
Transformace dat

TRANSFORMACE – CO TO JE?

- matematická funkce použitá na všechny původní hodnoty:

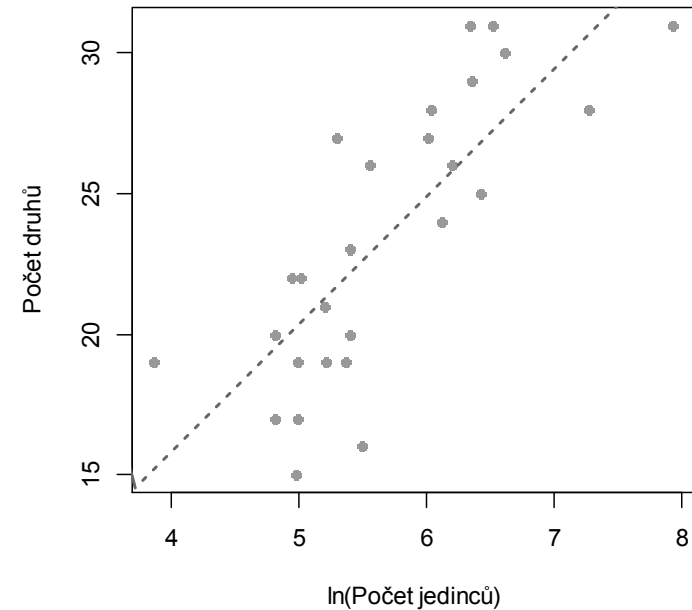
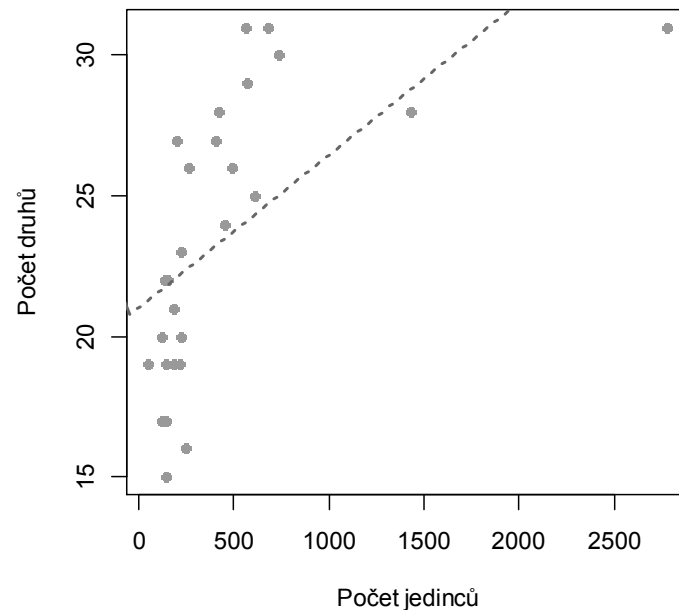
$$Y^* = f(Y)$$

- $f()$ – kontinuální, monotónická, většinou jednoduchá funkce
- nemění pořadí hodnot
- pořadí hodnot zůstane zachováno (transformace nemá vliv na neparametrické testy)
- mění relativní rozestupy mezi hodnotami a tudíž i varianci a tvar rozložení
- např. odmocnina, logaritmus



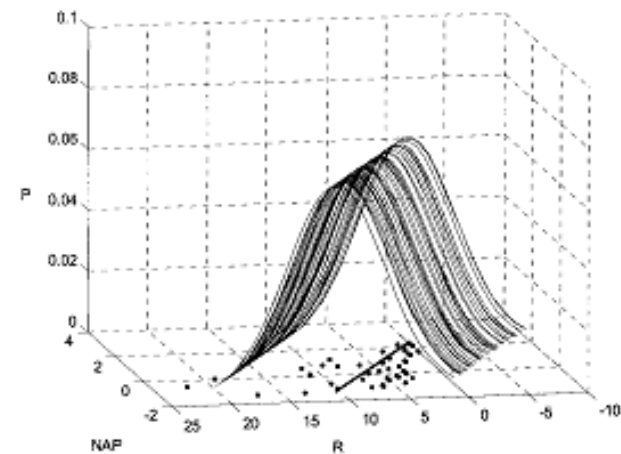
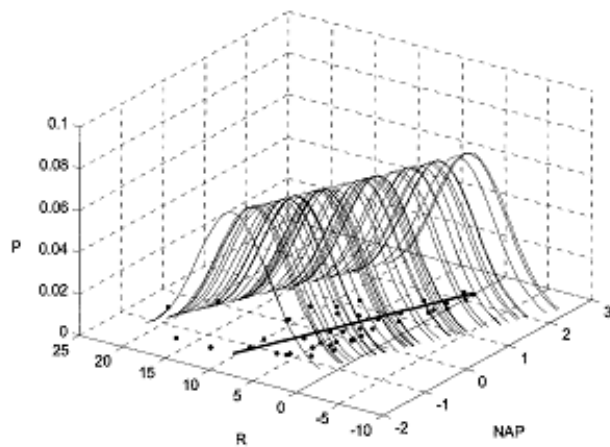
TRANSFORMACE – PROČ?

- nenormálně rozložená data
- heterogenní variance (heteroscedasticity)
- linearizace vztahů
 - lineární vztahy se lépe modelují a interpretují
- škála měření je arbitrární a nemusí odpovídat ekologickému významu proměnné

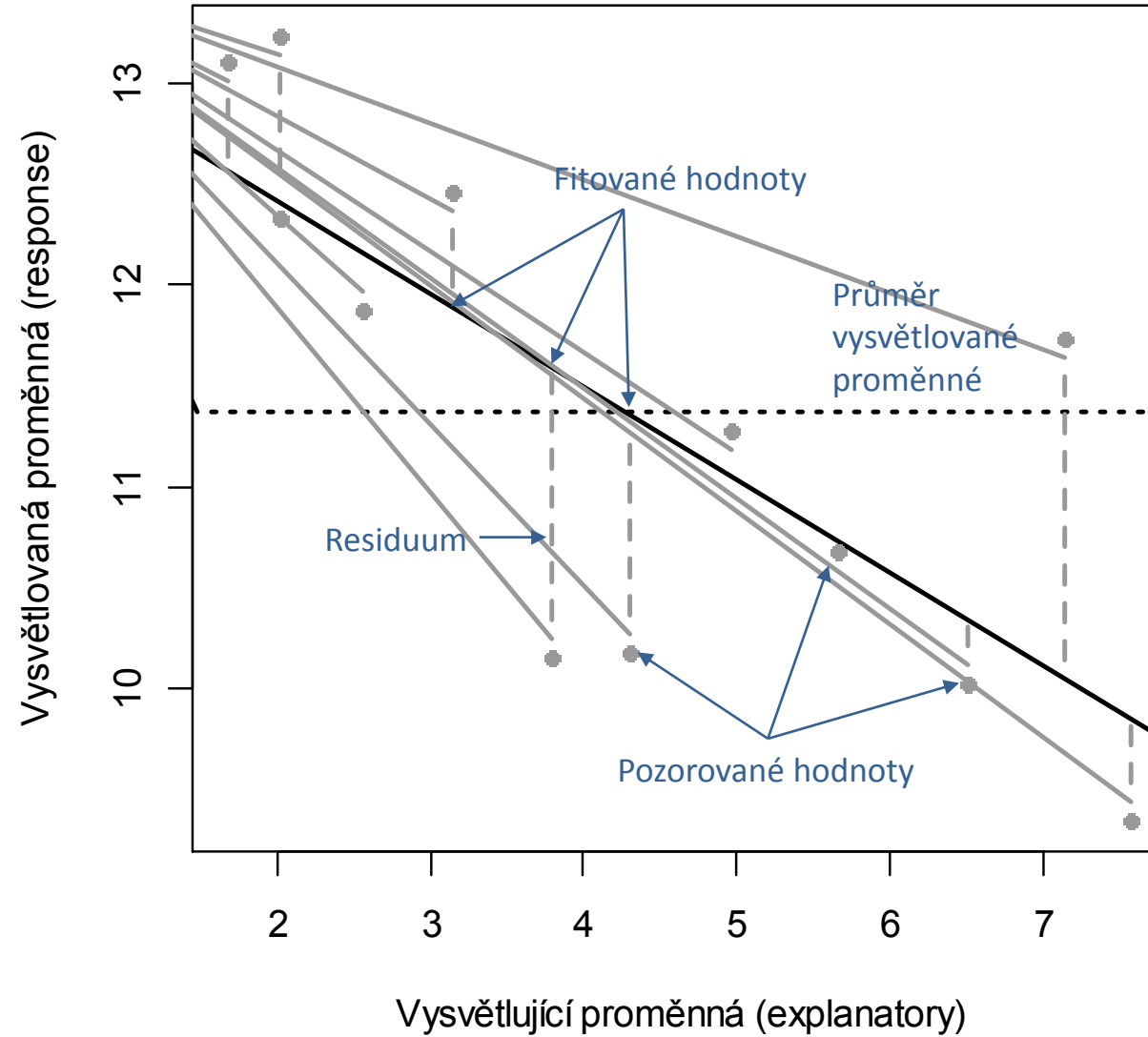


NORMALITA DAT

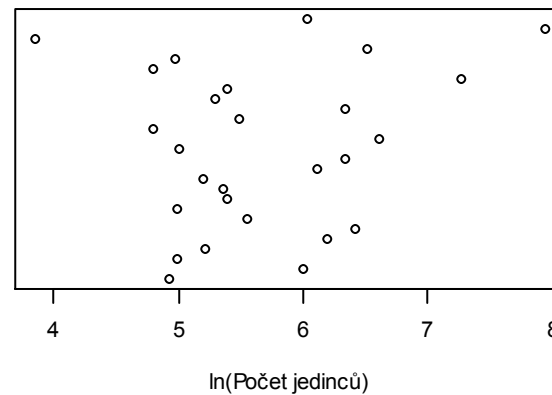
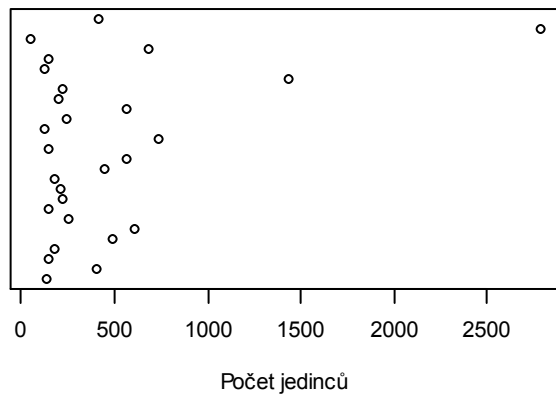
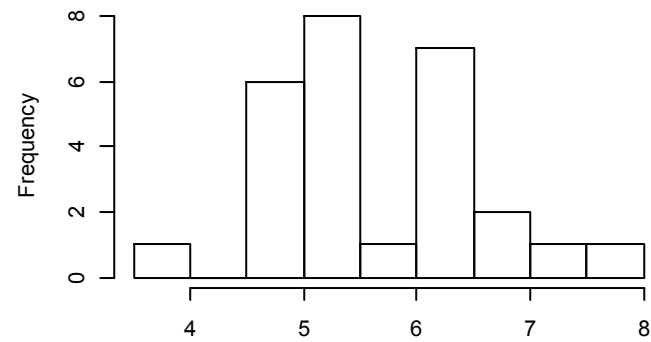
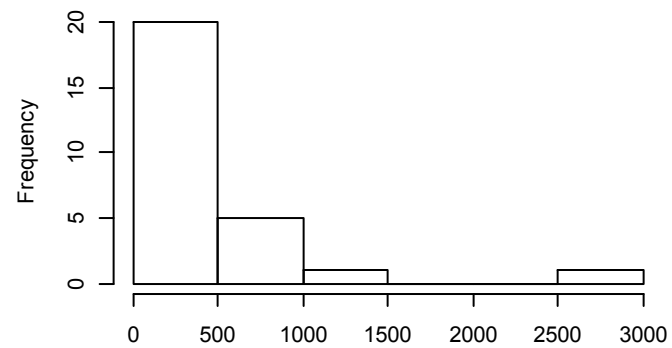
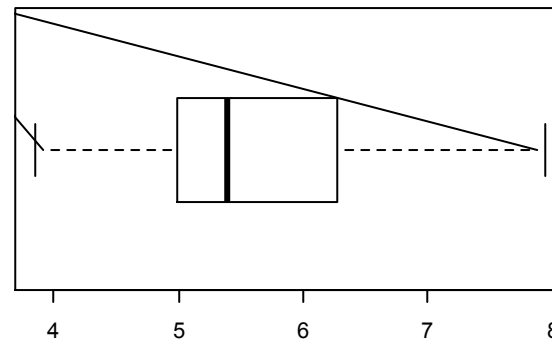
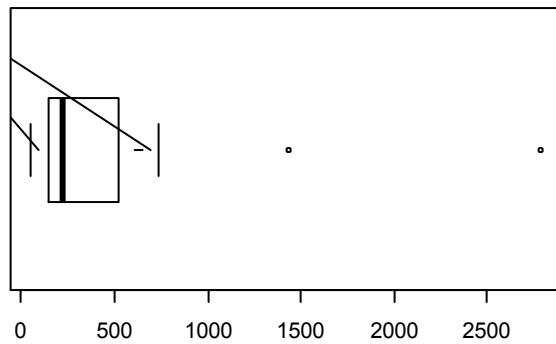
- mnohé testy hypotéz platné jen při splnění některých předpokladů
- jeden z nich je normalita rozložení **residuů**
- mylné a bezdůvodné testování normality prediktorů
- ideální prediktor má rozložení **uniformní**
 - četnost měření se nemění podél gradientu prediktoru



RESIDUA LINEÁRNÍHO MODELU

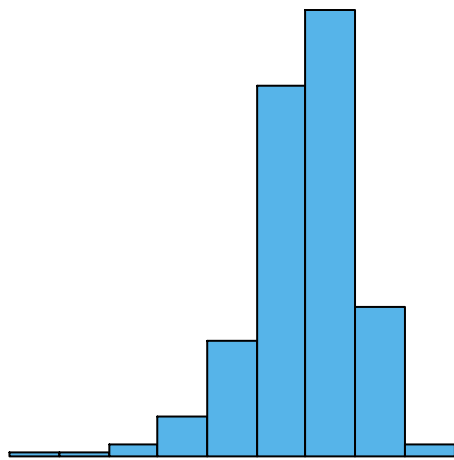


TRANSFORMACE

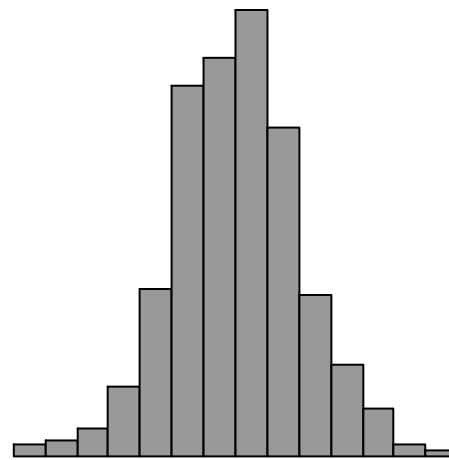


VÝBĚR TRANSFORMACE

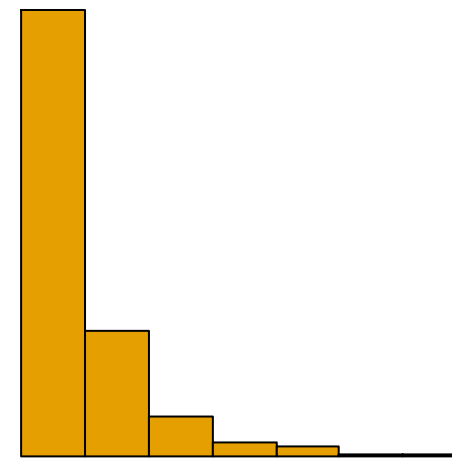
- tvar rozložení (sešikmenost – *skeweness*)
- vztah proměnných
- rozsah hodnot (zahrnují nulu nebo negativní hodnoty?)



Negativně (doleva)
sešikmené rozložení
(*left-skewed*)



Symetrické



pozitivně (doprava)
sešikmené rozložení
(*right-skewed*)

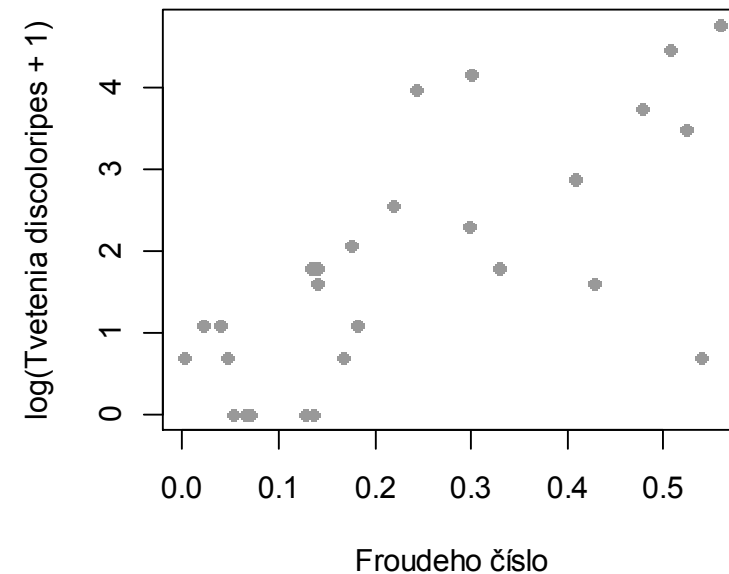
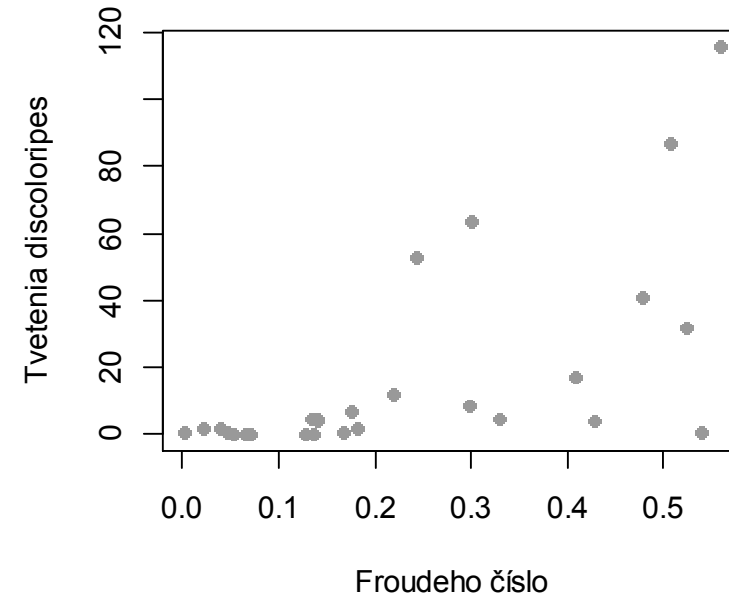
ČASTÉ TRANSFORMACE

- **Logaritmická transformace (*log transformation*)**
 - pro data s výrazně pozitivně (doprava) šikmou distribucí (*right skewed*)
 - Variabilita roste s hodnotami závislé proměnné
 - lognormální rozložení – běžné v ekologii
 - násobením sady nezávislých faktorů získáme lognormálně rozloženou proměnnou

$$Y^* = \log(Y) \quad \text{nebo} \quad Y^* = \log(a * Y + c)$$

$$y_{ij}^* = \log_{10}(y_{ij}) + 1 \quad \text{pro } y_{ij} > 0$$
$$y_{ij}^* = 0 \quad \text{pro } y_{ij} = 0$$

- na základě logaritmu nezáleží (10, 2, e)
 - Ale je třeba používat konzistentně
 - Pozor na zkratky ln/log vs. log/log₁₀ v češtině a angličtině
- konstanta $a > 1$ zabrání negativním hodnotám, pokud proměnná Y obsahuje hodnoty z intervalu $<0;1>$
- pokud proměnná obsahuje nuly (nebo negativní hodnoty), musíme přičíst konstantu c
- c by měla být stejného řádu jako měřené hodnoty (např. 0,01 při hodnotách od 0,00 do 0,09), u abundancí to odpovídá 1
- konstanta c má vliv na výsledné rozložení hodnot



ODMOCNINOVÁ A MOCNINOVÁ TRANSFORMACE

Odmocnina (*square root*)

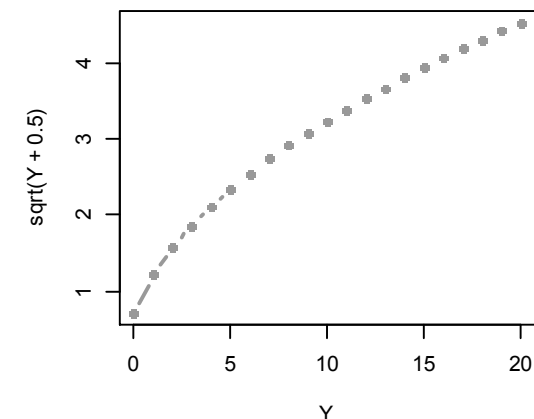
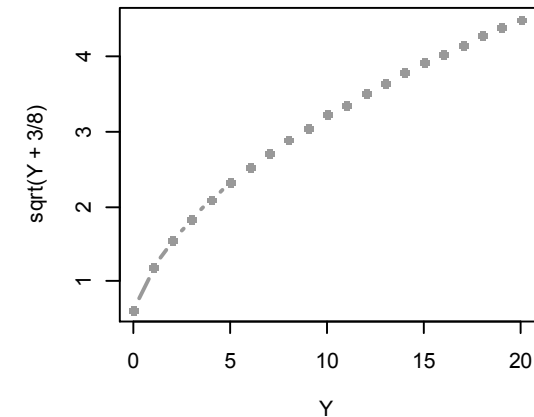
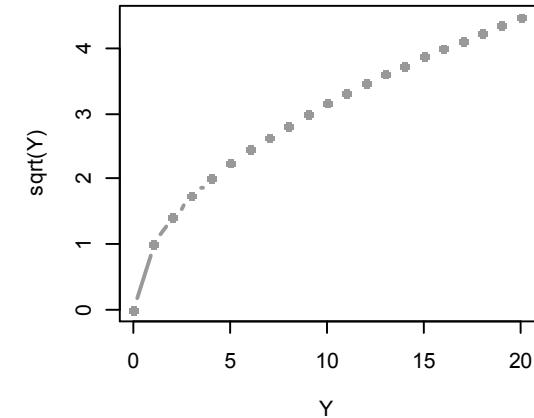
- na doprava sešikmené rozložení
- slabší efekt než logaritmus
- Pokryvnosti rostlin ve snímku – linearizace hodnot odvozených z velikosti plochy

případně

- pokud jsou v datech nuly, je někdy vhodné přidat konstantu c
- c např. 0,5 (Sokal & Rohlf, 1995) nebo $3/8$ (0,375) (Anscombe 1948)
- třetí a vyšší odmocnina je účinnější na více zešikmená data (čtvrtá odmocnina se používá pro abundance druhů s mnoha nulami a několika vysokými hodnotami)
- vysoká odmocnina se blíží logaritmu

Mocninná transformace (*power transformation*)

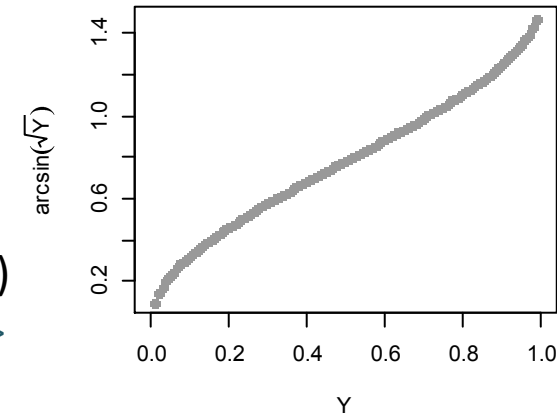
- vhodná pro data negativně (doleva) sešikmená (*left skewed*)



DALŠÍ TRANSFORMACE

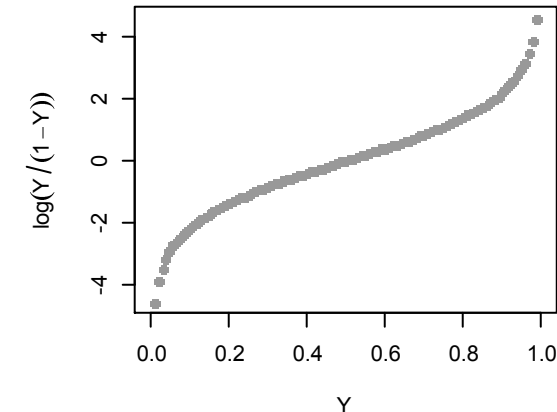
arcsin (angular transformation)

- vhodná pro procentické hodnoty (a obecně podíly)
 - použitelná pro hodnoty v intervalu $<-1; 1>$
 - jemně roztahuje hodnoty blízké 0 a 1



Logit

- vhodná pro podíly stejně jako arcsin
 - hodnoty od 0 do 1
 - roztahuje hodnoty blízké 0 a 1



Reciproká transformace (reciprocal transformation)

- vhodná pro poměry (například výška/hmotnost, počet dětí v populaci na počet žen atd.)
 - roztahuje hodnoty blízké nule
 - otáčí interpretaci

