M7222 – 3. cvičení : **GLM03b** (*Počty rostlinných druhů na pozemcích s různou biomasou a pH půdy*)

Máme k dispozici data, která se týkají dlouhodobého zemědělského experimentu. Bylo sledováno 90 pozemků (pastvin) o rozloze 25 m × 25 m, lišící se v biomase, pH půdy a druhové bohatosti (počet rostlinných druhů na celém pozemku). Je dobře známo, že s rostoucí biomasou dochází k poklesu druhové bohatosti. Ale zůstává otázka, zda rychlost poklesu nesouvisí s úrovní pH v půdě. Proto byly jednotlivé pozemky klasifikovány podle hodnoty pH v půdě do tří úrovní (nízká, střední a vysoká úroveň) a do experimentu bylo vybráno vždy po 30 pozemcích pro každou úroveň. Spojitá veličina `Biomass` je dlouhodobým průměrem naměřených červnových hodnot biomasy.

Nejprve načteme datový soubor pomocí příkazu `read.table()`, příkazem `str()` vypíšeme strukturu datového rámce.

```
> fileDat <- paste(data.library, "species.txt", sep = "")
> species <- read.table(fileDat, header = TRUE)
> str(species)
```

```
,data.frame,:   90 obs. of  3 variables:
 $ pH     : Factor w/ 3 levels "high","low","mid": 1 1 1 1 1 1 1 1 1 1 ...
 $ Biomass: num  0.469 1.731 2.09 3.926 4.367 ...
 $ Species: int  30 39 44 35 25 29 23 18 19 12 ...
```

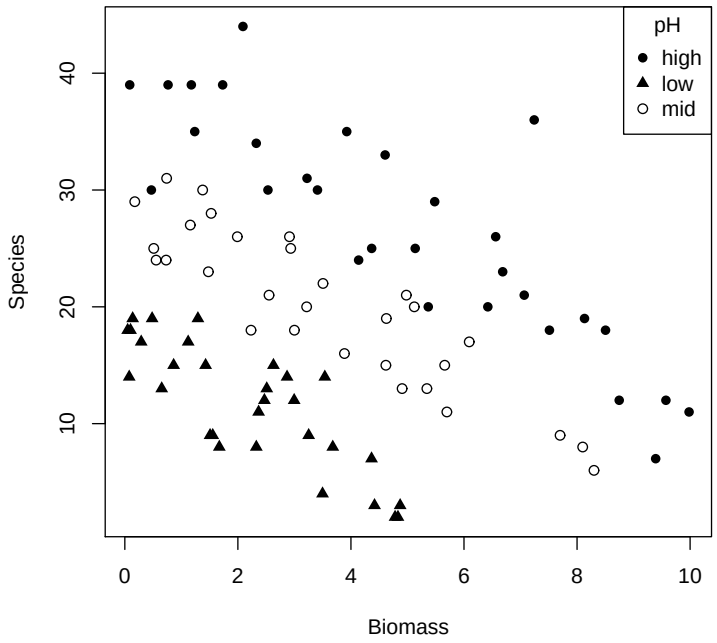
Pomocí příkazu `summary()` obdržíme první informace o datovém souboru.

```
> summary(species)
```

```
    pH        Biomass           Species
 high:30   Min.   :0.05018   Min.   : 2.00
 low :30   1st Qu.:1.44132   1st Qu.:12.25
 mid :30   Median :3.10836   Median :18.50
           Mean   :3.55701   Mean   :19.46
           3rd Qu.:5.08570   3rd Qu.:25.75
           Max.   :9.98177   Max.   :44.00
```

Abychom získali grafickou představu o datech, vykreslíme je pomocí příkazu `plot()` a pomocí různých symbolů odlišíme úrovně proměnné `pH`.

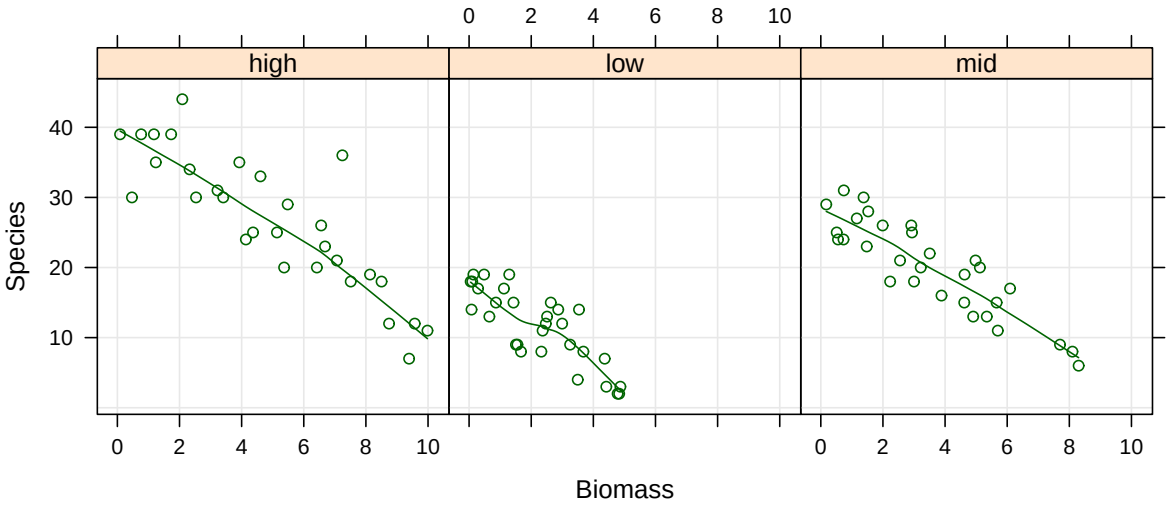
```
> PCH <- c(16, 17, 1)
> with(species, plot(Species ~ Biomass, pch = PCH[as.numeric(species$pH)]))
> legend("topright", with(species, levels(pH)), pch = PCH,
     title = "pH")
```



Obrázek 1: Bodový graf vstupních dat.

Chceme-li získat detailnější představu o tvaru trendových funkcí, použijeme `xyplot()` z knihovny `lattice`, kde díky volbě `"smooth"` jsou daty pomocí lokální polynomiální regresi (*loess*) proloženy odhady trendové funkce.

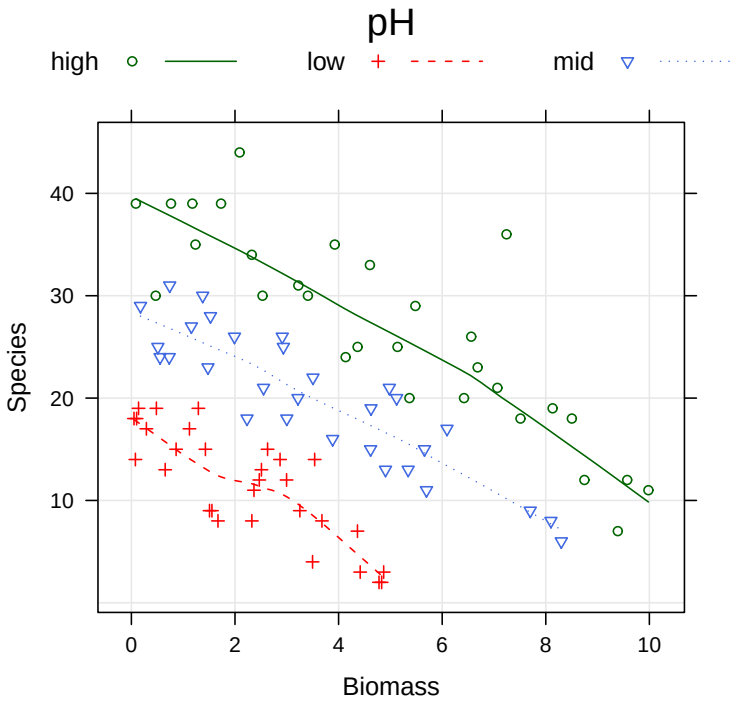
```
> print(xyplot(Species ~ Biomass | pH, data = species,
     type = list("g", "p", "smooth")))
```



Obrázek 2: Graf vstupních dat pomocí příkazu `xyplot()` do tří panelů.

Pokud chceme data znázornit společně v jediném panelu, napíšeme:

```
> print(xyplot(Species ~ Biomass, groups = pH, data = species,
     type = list("g", "p", "smooth"), auto.key = list(points = TRUE,
         lines = TRUE, columns = 3, title = "pH")))
```



Obrázek 3: Graf vstupních dat pomocí příkazu `xyplot()` do jediného panelu.

Vzhledem k tomu, že závisle proměnná `Species` značí počet rostlinných druhů, budeme předpokládat, že její rozdělení je Poissonovo. Pro Poissonovo rozdělení máme k dispozici tři linkovací funkce

- logaritmus (kanonická linkovací funkce),
- odmocnina,
- identická funkce.

Z předchozích grafů je patrné, že při modelování vztahu mezi poissonovskou proměnnou `Species` a proměnnou `Biomass` (pro tři typy půd) můžeme vyzkoušet všechny tři linkovací funkce

Protože datový soubor obsahuje také kategoriální proměnnou `pH`, půjde o analýzu kovariance s lineárním prediktorem tvaru

$$\eta_{ij} = \beta_0 + \alpha_{0,j} + (\beta_1 + \alpha_{1,j})x_{ij} \qquad i = 1, \ldots, 30 \qquad j = 2, 3.$$

Pomocí příkazu `glm()` obdržíme odhady parametrů modelu postupně pro jednotlivé linkovací funkce.

```
> m1 <- glm(Species ~ Biomass * pH, family = poisson(log),
     data = species)
> summary(m1)
```

```
Call:
glm(formula = Species ~ Biomass * pH, family = poisson(log),
    data = species)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-2.49779  -0.74845  -0.04023   0.55745   3.22975

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)   3.76812    0.06153  61.240  < 2e-16 ***
Biomass      -0.10713    0.01249  -8.577  < 2e-16 ***
pHlow        -0.81557    0.10284  -7.931 2.18e-15 ***
pHmid        -0.33146    0.09217  -3.596 0.000323 ***
Biomass:pHlow -0.15503    0.04003  -3.873 0.000108 ***
Biomass:pHmid -0.03189    0.02308  -1.382 0.166954
---
Signif. codes:  0 ,***, 0.001 ,**, 0.01 ,*, 0.05 ,., 0.1 , , 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 452.346  on 89  degrees of freedom
Residual deviance:  83.201  on 84  degrees of freedom
AIC: 514.39

Number of Fisher Scoring iterations: 4

> m2 <- glm(Species ~ Biomass * pH, family = poisson(identity),
     data = species)
> summary(m2)


Call:
glm(formula = Species ~ Biomass * pH, family = poisson(identity),
    data = species)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-1.96538  -0.66136  -0.05127   0.47995   3.22838

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)   41.4276     2.0318  20.390  < 2e-16 ***
Biomass       -2.9675     0.3050  -9.731  < 2e-16 ***
pHlow        -22.9839     2.3597  -9.740  < 2e-16 ***
pHmid        -12.1627     2.5642  -4.743 2.10e-06 ***
Biomass:pHlow  -0.1288     0.4624  -0.278    0.781
Biomass:pHmid   0.3360     0.4322   0.777    0.437
---
Signif. codes:  0 ,***, 0.001 ,**, 0.01 ,*, 0.05 ,., 0.1 , , 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 452.35  on 89  degrees of freedom
Residual deviance:  67.07  on 84  degrees of freedom
AIC: 498.26

Number of Fisher Scoring iterations: 6
```

```
> m3 <- glm(Species ~ Biomass * pH, family = poisson(sqrt),
     data = species)
> summary(m3)


Call:
glm(formula = Species ~ Biomass * pH, family = poisson(sqrt),
    data = species)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.2430  -0.6885  -0.1093   0.5408   3.2633

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)   6.51567    0.17901  36.399  < 2e-16 ***
Biomass      -0.28531    0.03124  -9.133  < 2e-16 ***
pHlow        -2.16972    0.24037  -9.026  < 2e-16 ***
pHmid        -1.01278    0.24477  -4.138 3.51e-05 ***
Biomass:pHlow -0.17341    0.06711  -2.584  0.00977 **
Biomass:pHmid -0.02357    0.05052  -0.467  0.64075
---
Signif. codes:  0 ,***, 0.001 ,**, 0.01 ,*, 0.05 ,., 0.1 , , 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 452.346  on 89  degrees of freedom
Residual deviance:  74.876  on 84  degrees of freedom
AIC: 506.07

Number of Fisher Scoring iterations: 5
```

Všechny tři postupy ukazují, že jednotlivé kovariáty (až na poslední) jsou statisticky významné. Vhodnost modelu ověříme pomocí Waldova testu.

```
> library(lmtest)
> waldtest(m1, test = "Chisq")


Wald test

Model 1: Species ~ Biomass * pH
Model 2: Species ~ 1
  Res.Df Df  Chisq Pr(>Chisq)
1     84
2     89 -5 333.95  < 2.2e-16 ***
---
Signif. codes:  0 ,***, 0.001 ,**, 0.01 ,*, 0.05 ,., 0.1 , , 1


> waldtest(m2, test = "Chisq")


Wald test

Model 1: Species ~ Biomass * pH
```

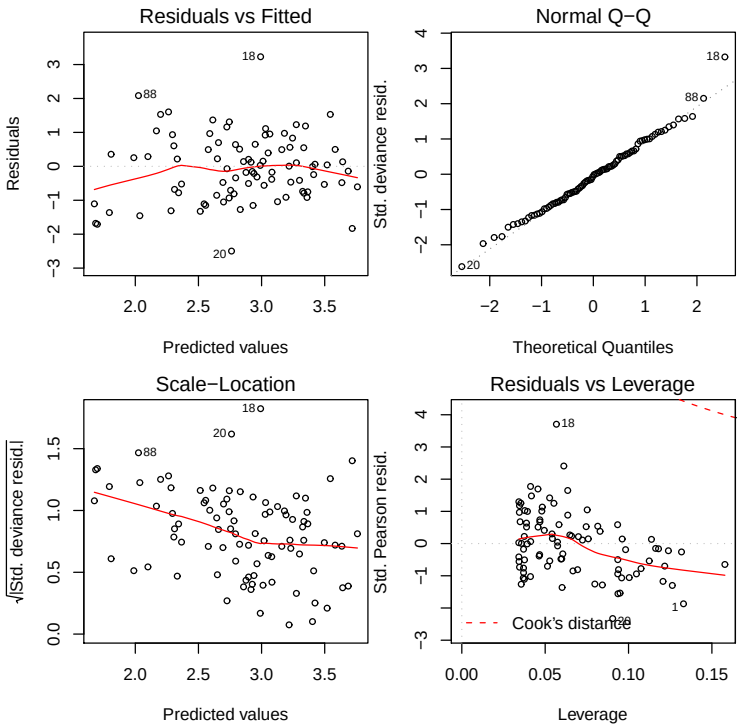
```
Model 2: Species ~ 1
  Res.Df Df  Chisq Pr(>Chisq)
1     84
2     89 -5 473.87  < 2.2e-16 ***
---
Signif. codes:  0 ,***, 0.001 ,**, 0.01 ,*, 0.05 ,., 0.1 , , 1
```

```
> waldtest(m3, test = "Chisq")
```

```
Wald test

Model 1: Species ~ Biomass * pH
Model 2: Species ~ 1
  Res.Df Df  Chisq Pr(>Chisq)
1     84
2     89 -5 397.29  < 2.2e-16 ***
---
Signif. codes:  0 ,***, 0.001 ,**, 0.01 ,*, 0.05 ,., 0.1 , , 1
```

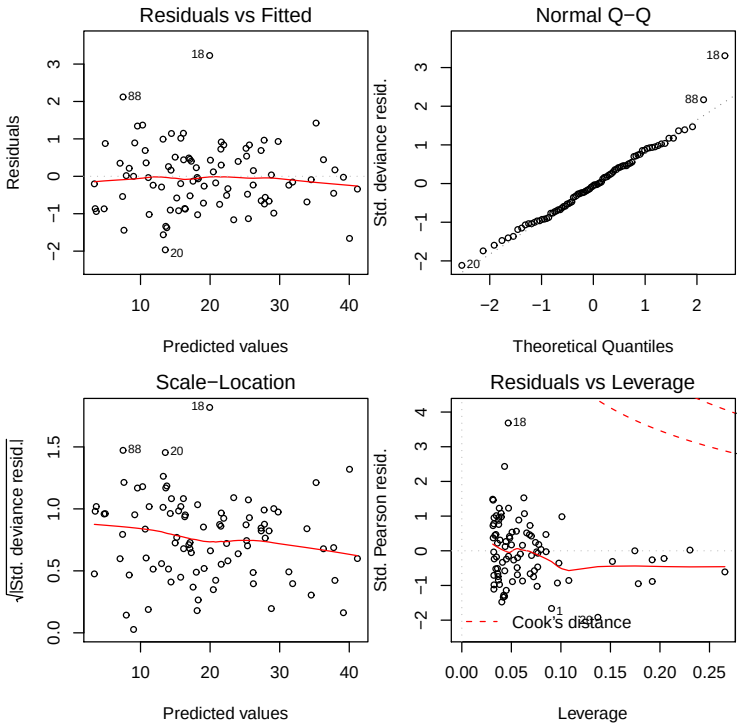
Z výsledku je zřejmé, že oproti nulovému modelu se přidáním kovariát `Biomass` a `pH` model výrazně zlepšil. Nesmíme zapomenout také na analýzu reziduí.

```
> par(mfrow = c(2, 2), mar = c(4, 4, 2, 0) + 0.05)
> plot(m1, cex = 0.75)
```



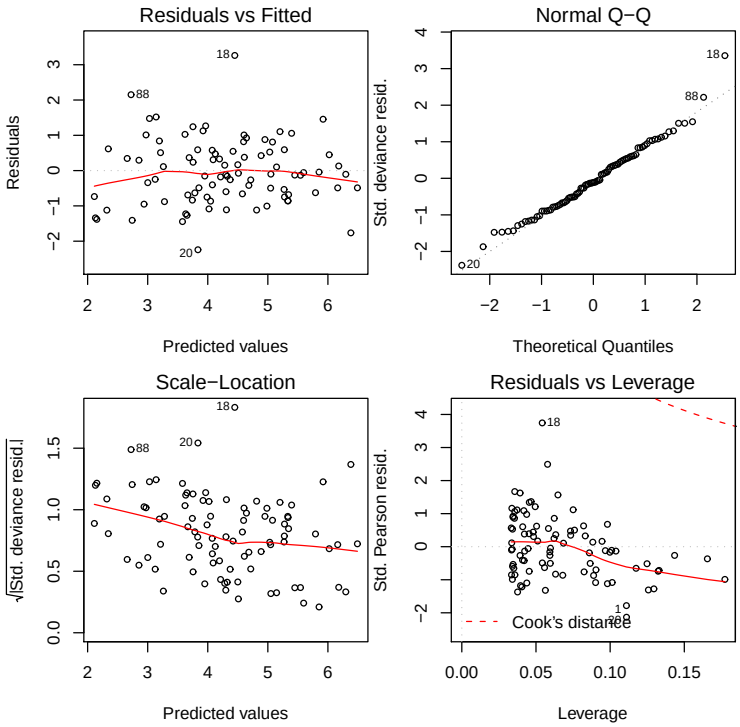
Obrázek 4: Analýza reziduí (různé průsečíky i směrnice) pro kanonickou linkovací funkci.

```
> par(mfrow = c(2, 2), mar = c(4, 4, 2, 0) + 0.05)
> plot(m2, cex = 0.75)
```



Obrázek 5: Analýza reziduí (různé průsečíky i směrnice) pro identickou linkovací funkci.

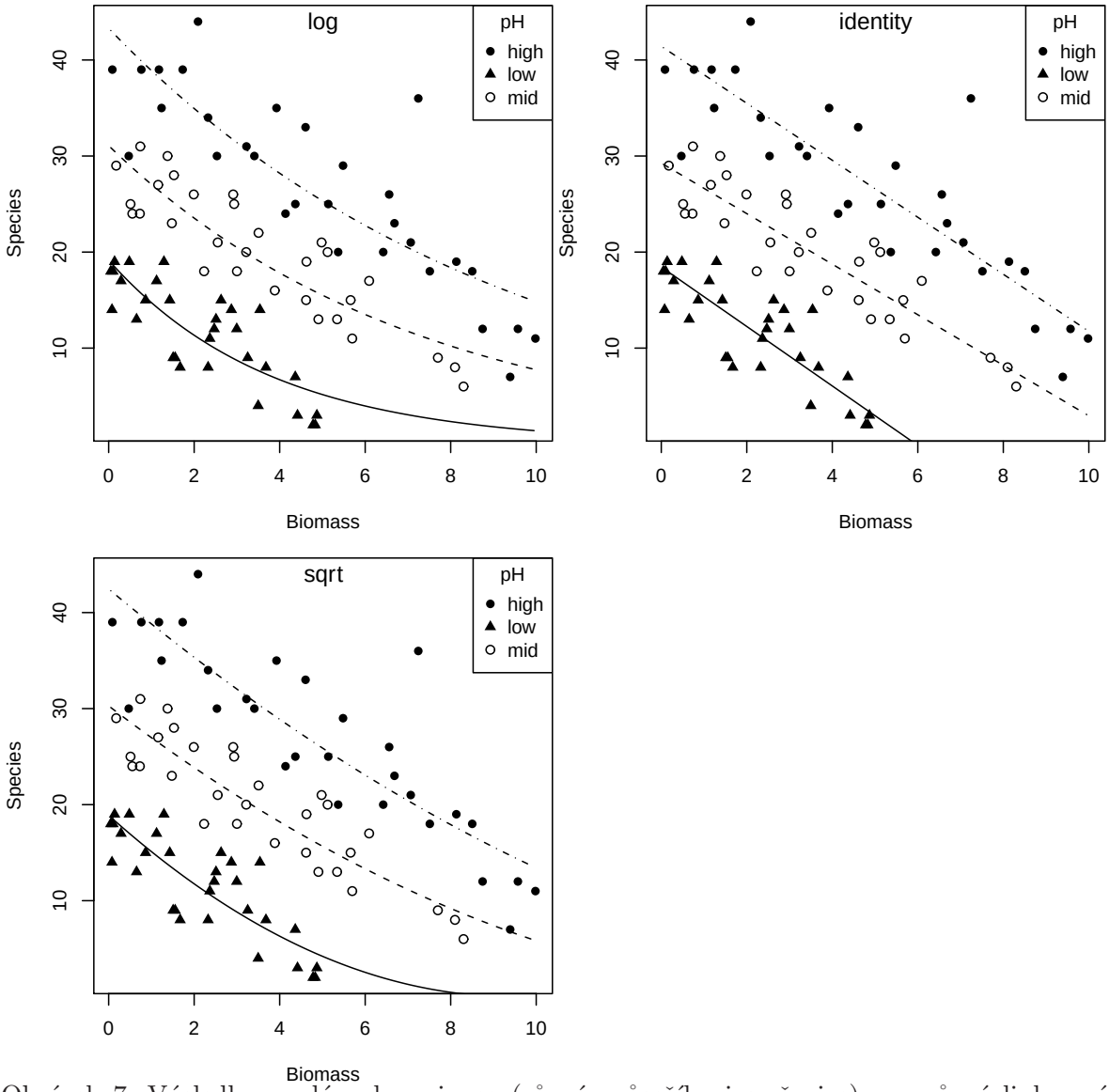
```
> par(mfrow = c(2, 2), mar = c(4, 4, 2, 0) + 0.05)
> plot(m3, cex = 0.75)
```



Obrázek 6: Analýza reziduí (různé průsečíky i směrnice) pro odmocninovou linkovací funkci.

Výsledné odhady křivek pro všechny tři volby linkovací funkce graficky znázorníme.

```
> ab <- with(species, range(Biomass))
> xv = seq(ab[1], ab[2], 0.1)
> par(mfrow = c(2, 2), mar = c(4, 4, 1, 0) + 0.05)
> with(species, plot(Species ~ Biomass, pch = PCH[as.numeric(species$pH)]))
> phv = factor(rep("low", length(xv)))
> yv = predict(m1, list(pH = phv, Biomass = xv), type = "response")
> lines(xv, yv, lty = 1)
> phv = factor(rep("mid", length(xv)))
> yv = predict(m1, list(pH = phv, Biomass = xv), type = "response")
> lines(xv, yv, lty = 2)
> phv = factor(rep("high", length(xv)))
> yv = predict(m1, list(pH = phv, Biomass = xv), type = "response")
> lines(xv, yv, lty = 4)
> legend("topright", with(species, levels(pH)), pch = PCH,
    title = "pH")
> mtext("log", line = -1.25)
> with(species, plot(Species ~ Biomass, pch = PCH[as.numeric(species$pH)]))
> phv = factor(rep("low", length(xv)))
> yv = predict(m2, list(pH = phv, Biomass = xv), type = "response")
> lines(xv, yv, lty = 1)
> phv = factor(rep("mid", length(xv)))
> yv = predict(m2, list(pH = phv, Biomass = xv), type = "response")
> lines(xv, yv, lty = 2)
> phv = factor(rep("high", length(xv)))
> yv = predict(m2, list(pH = phv, Biomass = xv), type = "response")
> lines(xv, yv, lty = 4)
> legend("topright", with(species, levels(pH)), pch = PCH,
    title = "pH")
> mtext("identity", line = -1.25)
> with(species, plot(Species ~ Biomass, pch = PCH[as.numeric(species$pH)]))
> phv = factor(rep("low", length(xv)))
> yv = predict(m3, list(pH = phv, Biomass = xv), type = "response")
> lines(xv, yv, lty = 1)
> phv = factor(rep("mid", length(xv)))
> yv = predict(m3, list(pH = phv, Biomass = xv), type = "response")
> lines(xv, yv, lty = 2)
> phv = factor(rep("high", length(xv)))
> yv = predict(m3, list(pH = phv, Biomass = xv), type = "response")
> lines(xv, yv, lty = 4)
> legend("topright", with(species, levels(pH)), pch = PCH,
    title = "pH")
> mtext("sqrt", line = -1.25)
```



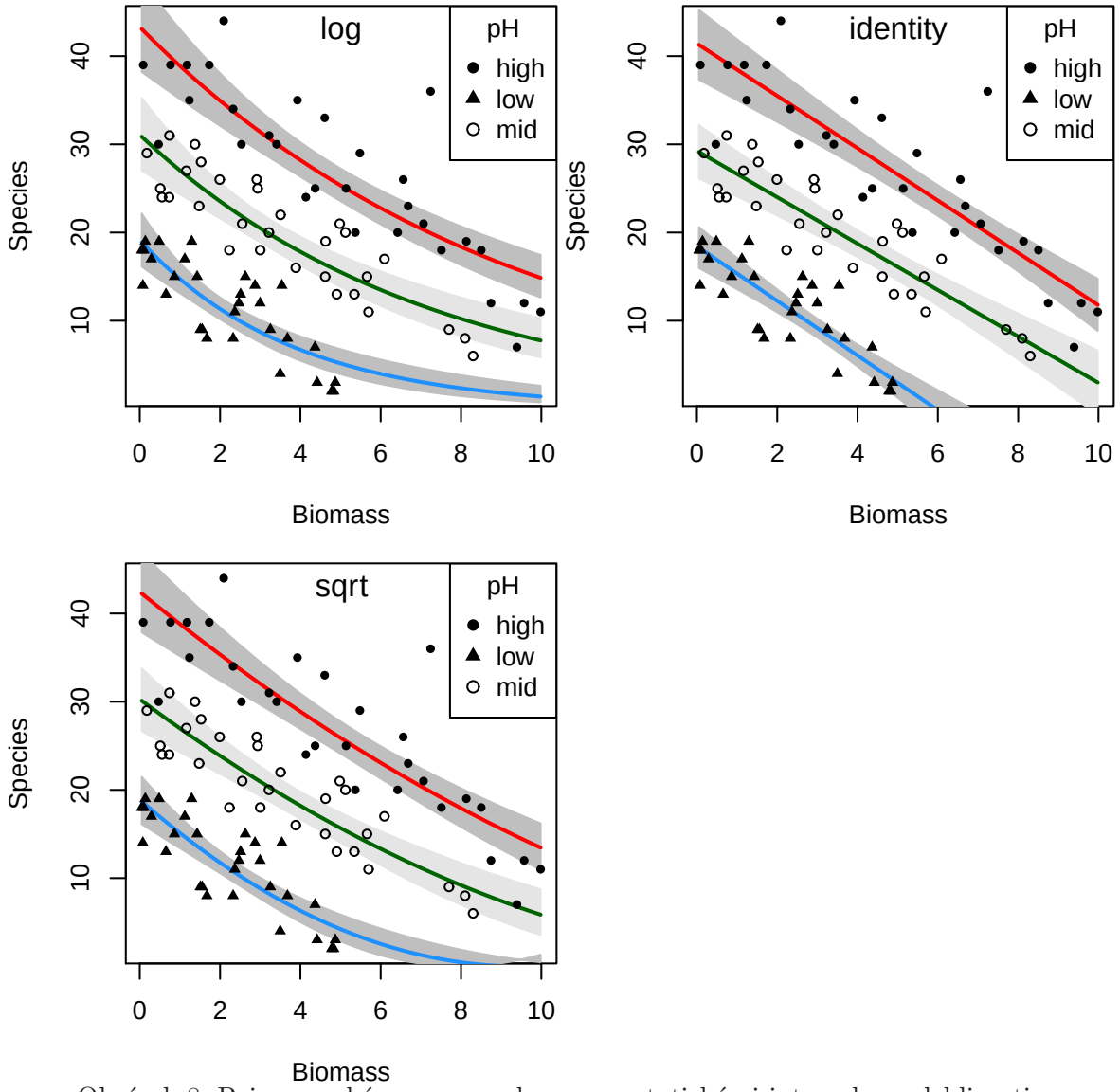
Obrázek 7: Výsledky analýzy kovariance (různé průsečíky i směrnice) pro různé linkovací funkce.

Nyní ještě vytvoříme obdobné grafy s intervaly spolehlivosti kolem střední hodnoty.

```
> par(mfrow = c(2, 2), mar = c(4, 4, 1, 0) + 0.05)
> with(species, plot(Species ~ Biomass, type = "n"))
> ab <- with(species, range(Biomass))
> nn <- 200
> Levels <- with(species, levels(pH))
> Fac1 <- factor(rep(Levels[1], nn), levels = Levels)
> Fac2 <- factor(rep(Levels[2], nn), levels = Levels)
> Fac3 <- factor(rep(Levels[3], nn), levels = Levels)
> xx <- seq(ab[1], ab[2], length.out = nn)
> yy1 <- predict(m1, list(Biomass = xx, pH = Fac1), type = "response")
> yy2 <- predict(m1, list(Biomass = xx, pH = Fac2), type = "response")
> yy3 <- predict(m1, list(Biomass = xx, pH = Fac3), type = "response")
```

```
> Predicted <- predict(m1, list(Biomass = xx, pH = Fac1),
     type = "link", se = T)
> CI.L.log <- exp(Predicted$fit - 1.96 * Predicted$se.fit)
> CI.H.log <- exp(Predicted$fit + 1.96 * Predicted$se.fit)
> x <- c(xx, rev(xx))
> y <- c(CI.L.log, rev(CI.H.log))
> polygon(x, y, col = "gray75", border = "gray75")
> Predicted <- predict(m1, list(Biomass = xx, pH = Fac2),
     type = "link", se = T)
> CI.L.log <- exp(Predicted$fit - 1.96 * Predicted$se.fit)
> CI.H.log <- exp(Predicted$fit + 1.96 * Predicted$se.fit)
> x <- c(xx, rev(xx))
> y <- c(CI.L.log, rev(CI.H.log))
> polygon(x, y, col = "gray75", border = "gray75")
> Predicted <- predict(m1, list(Biomass = xx, pH = Fac3),
     type = "link", se = T)
> CI.L.log <- exp(Predicted$fit - 1.96 * Predicted$se.fit)
> CI.H.log <- exp(Predicted$fit + 1.96 * Predicted$se.fit)
> x <- c(xx, rev(xx))
> y <- c(CI.L.log, rev(CI.H.log))
> polygon(x, y, col = "gray90", border = "gray90")
> lines(xx, yy1, col = "red", lwd = 2)
> lines(xx, yy2, col = "dodgerblue", lwd = 2)
> lines(xx, yy3, col = "darkgreen", lwd = 2)
> with(species, points(Species ~ Biomass, cex = 0.75, pch = PCH[as.numeric(species$pH)]))
> legend("topright", with(species, levels(pH)), pch = PCH,
     title = "pH")
> mtext("log", line = -1.25)
> with(species, plot(Species ~ Biomass, type = "n"))
> yy1 <- predict(m2, list(Biomass = xx, pH = Fac1), type = "response")
> yy2 <- predict(m2, list(Biomass = xx, pH = Fac2), type = "response")
> yy3 <- predict(m2, list(Biomass = xx, pH = Fac3), type = "response")
> Predicted <- predict(m2, list(Biomass = xx, pH = Fac1),
     type = "link", se = T)
> CI.L.log <- (Predicted$fit - 1.96 * Predicted$se.fit)
> CI.H.log <- (Predicted$fit + 1.96 * Predicted$se.fit)
> x <- c(xx, rev(xx))
> y <- c(CI.L.log, rev(CI.H.log))
> polygon(x, y, col = "gray75", border = "gray75")
> Predicted <- predict(m2, list(Biomass = xx, pH = Fac2),
     type = "link", se = T)
> CI.L.log <- (Predicted$fit - 1.96 * Predicted$se.fit)
> CI.H.log <- (Predicted$fit + 1.96 * Predicted$se.fit)
> x <- c(xx, rev(xx))
> y <- c(CI.L.log, rev(CI.H.log))
> polygon(x, y, col = "gray75", border = "gray75")
> Predicted <- predict(m2, list(Biomass = xx, pH = Fac3),
     type = "link", se = T)
> CI.L.log <- (Predicted$fit - 1.96 * Predicted$se.fit)
> CI.H.log <- (Predicted$fit + 1.96 * Predicted$se.fit)
> x <- c(xx, rev(xx))
> y <- c(CI.L.log, rev(CI.H.log))
> polygon(x, y, col = "gray90", border = "gray90")
> lines(xx, yy1, col = "red", lwd = 2)
> lines(xx, yy2, col = "dodgerblue", lwd = 2)
> lines(xx, yy3, col = "darkgreen", lwd = 2)
> with(species, points(Species ~ Biomass, cex = 0.75, pch = PCH[as.numeric(species$pH)]))
```

```
> legend("topright", with(species, levels(pH)), pch = PCH,
    title = "pH")
> mtext("identity", line = -1.25)
> with(species, plot(Species ~ Biomass, type = "n"))
> yy1 <- predict(m3, list(Biomass = xx, pH = Fac1), type = "response")
> yy2 <- predict(m3, list(Biomass = xx, pH = Fac2), type = "response")
> yy3 <- predict(m3, list(Biomass = xx, pH = Fac3), type = "response")
> Predicted <- predict(m3, list(Biomass = xx, pH = Fac1),
    type = "link", se = T)
> CI.L.log <- (Predicted$fit - 1.96 * Predicted$se.fit)^2
> CI.H.log <- (Predicted$fit + 1.96 * Predicted$se.fit)^2
> x <- c(xx, rev(xx))
> y <- c(CI.L.log, rev(CI.H.log))
> polygon(x, y, col = "gray75", border = "gray75")
> Predicted <- predict(m3, list(Biomass = xx, pH = Fac2),
    type = "link", se = T)
> CI.L.log <- (Predicted$fit - 1.96 * Predicted$se.fit)^2
> CI.H.log <- (Predicted$fit + 1.96 * Predicted$se.fit)^2
> x <- c(xx, rev(xx))
> y <- c(CI.L.log, rev(CI.H.log))
> polygon(x, y, col = "gray75", border = "gray75")
> Predicted <- predict(m3, list(Biomass = xx, pH = Fac3),
    type = "link", se = T)
> CI.L.log <- (Predicted$fit - 1.96 * Predicted$se.fit)^2
> CI.H.log <- (Predicted$fit + 1.96 * Predicted$se.fit)^2
> x <- c(xx, rev(xx))
> y <- c(CI.L.log, rev(CI.H.log))
> polygon(x, y, col = "gray90", border = "gray90")
> lines(xx, yy1, col = "red", lwd = 2)
> lines(xx, yy2, col = "dodgerblue", lwd = 2)
> lines(xx, yy3, col = "darkgreen", lwd = 2)
> with(species, points(Species ~ Biomass, cex = 0.75, pch = PCH[as.numeric(species$pH)]))
> legend("topright", with(species, levels(pH)), pch = PCH,
    title = "pH")
> mtext("sqrt", line = -1.25)
```



Obrázek 8: Poissonovská regrese spolu s asymptotickými intervaly spolehlivosti.

Nyní zkusíme model zjednodušit, a to tak, že budeme předpokládat, že linerární prediktor má pro jednotlivé druhy pH různé průsečíky, ale směrnice zůstává pro všechny tři typy stejná.

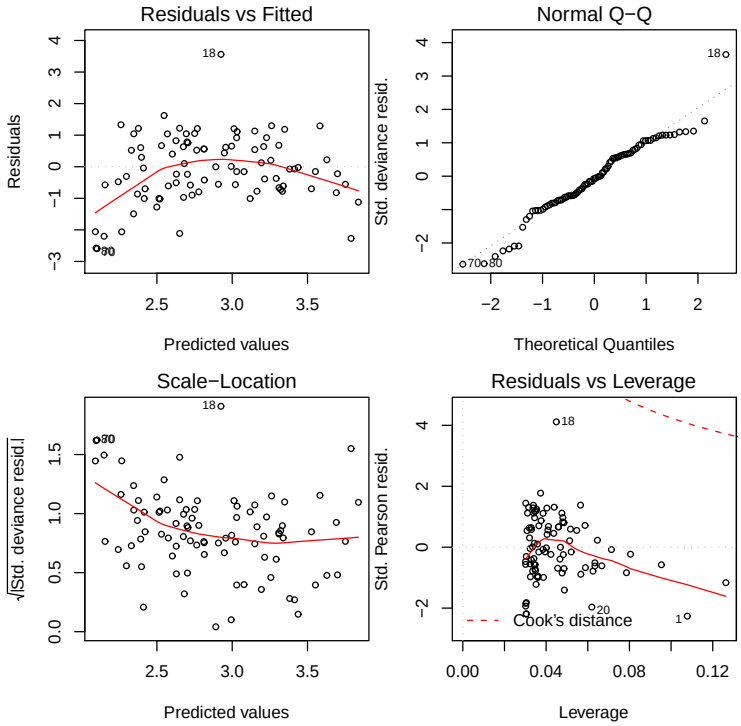
```
> m1b <- update(m1, ~. - pH:Biomass)
> summary(m1b)


Call:
glm(formula = Species ~ Biomass + pH, family = poisson(log),
    data = species)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-2.59586  -0.69887  -0.07373   0.66472   3.56040
```

```
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  3.84894    0.05281  72.885  < 2e-16 ***
Biomass     -0.12756    0.01014 -12.579  < 2e-16 ***
pHlow       -1.13639    0.06720 -16.910  < 2e-16 ***
pHmid       -0.44516    0.05486  -8.114 4.88e-16 ***
---
Signif. codes:  0 ,***, 0.001 ,**, 0.01 ,*, 0.05 ,., 0.1 , , 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 452.346  on 89  degrees of freedom
Residual deviance:  99.242  on 86  degrees of freedom
AIC: 526.43

Number of Fisher Scoring iterations: 4


> m2b <- update(m2, ~. - pH:Biomass)
> summary(m2b)


Call:
glm(formula = Species ~ Biomass + pH, family = poisson(identity),
    data = species)

Deviance Residuals:
      Min         1Q     Median         3Q        Max
-2.048099  -0.714188  -0.002176   0.498606   3.195710

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   40.860      1.407  29.037   <2e-16 ***
Biomass       -2.871      0.183 -15.684   <2e-16 ***
pHlow        -23.068      1.180 -19.548   <2e-16 ***
pHmid        -10.508      1.195  -8.791   <2e-16 ***
---
Signif. codes:  0 ,***, 0.001 ,**, 0.01 ,*, 0.05 ,., 0.1 , , 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 452.346  on 89  degrees of freedom
Residual deviance:  68.126  on 86  degrees of freedom
AIC: 495.32

Number of Fisher Scoring iterations: 6


> m3b <- update(m3, ~. - pH:Biomass)
> summary(m3b)


Call:
glm(formula = Species ~ Biomass + pH, family = poisson(sqrt),
    data = species)

Deviance Residuals:
```

```
    Min       1Q   Median       3Q      Max
-2.0358  -0.7273  -0.0930   0.5691   3.4287

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  6.67369    0.14436   46.23   <2e-16 ***
Biomass     -0.31733    0.02269  -13.99   <2e-16 ***
pHlow       -2.63199    0.14298  -18.41   <2e-16 ***
pHmid       -1.14113    0.13300   -8.58   <2e-16 ***
---
Signif. codes:  0 ,***, 0.001 ,**, 0.01 ,*, 0.05 ,., 0.1 , , 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 452.346  on 89  degrees of freedom
Residual deviance:  81.133  on 86  degrees of freedom
AIC: 508.32

Number of Fisher Scoring iterations: 5
```

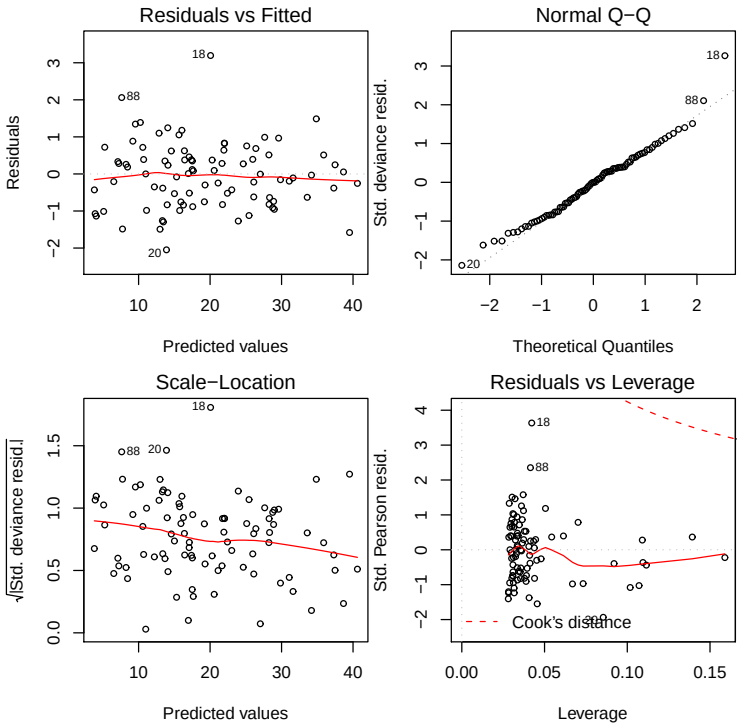
Vidíme, že všechny regresory jsou statisticky významné. Nesmíme zapomenout také na analýzu reziduí.

```
> par(mfrow = c(2, 2), mar = c(4, 4, 2, 0) + 0.05)
> plot(m1b, cex = 0.75)
```



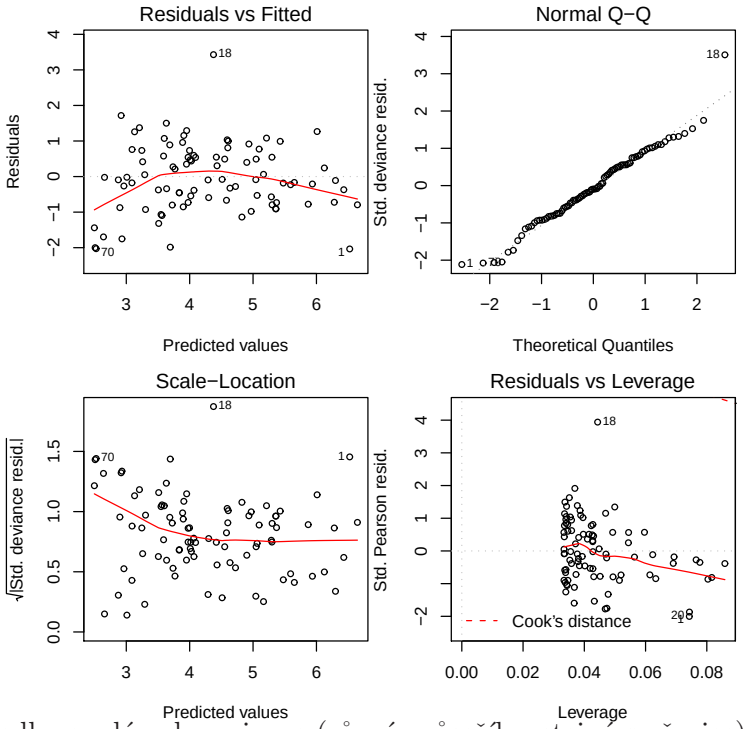
Obrázek 9: Analýza reziduí (různé průsečíky, stejná směrnice) pro kanonickou linkovací funkci.

```
> par(mfrow = c(2, 2), mar = c(4, 4, 2, 0) + 0.05)
> plot(m2b, cex = 0.75)
```



Obrázek 10: Analýza reziduí (různé průsečíky, stejná směrnice) pro identickou linkovací funkci.

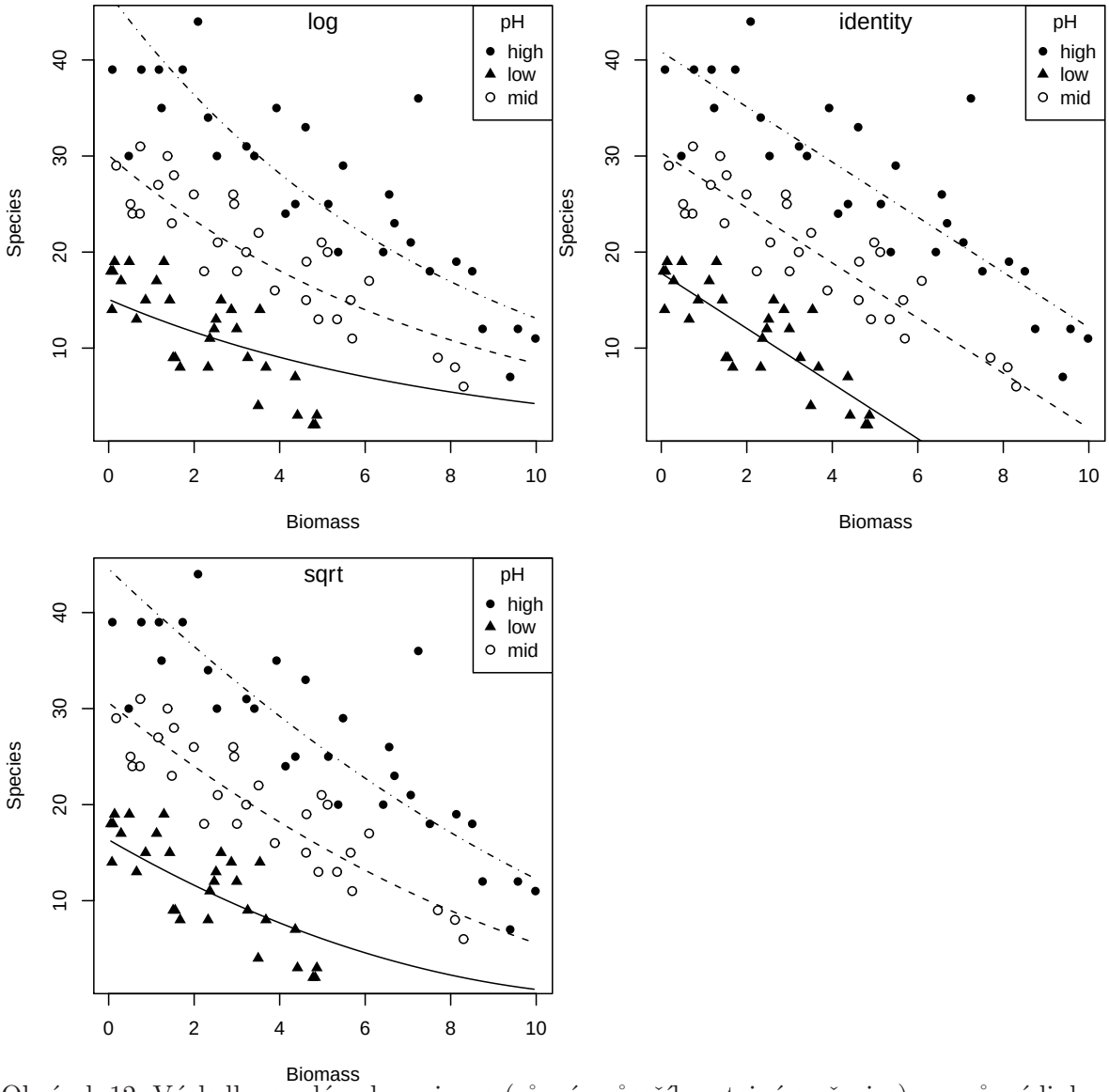
```
> par(mfrow = c(2, 2), mar = c(4, 4, 2, 0) + 0.05)
> plot(m3b, cex = 0.75)
```



Obrázek 11: Výsledky analýzy kovariance (různé průsečíky, stejná směrnice) pro různé linkovací funkce.

Výsledné odhady křivek pro všechny tři volby linkovací funkce graficky znázorníme.

```
> ab <- with(species, range(Biomass))
> xv = seq(ab[1], ab[2], 0.1)
> par(mfrow = c(2, 2), mar = c(4, 4, 1, 0) + 0.05)
> with(species, plot(Species ~ Biomass, pch = PCH[as.numeric(species$pH)]))
> phv = factor(rep("low", length(xv)))
> yv = predict(m1b, list(pH = phv, Biomass = xv), type = "response")
> lines(xv, yv, lty = 1)
> phv = factor(rep("mid", length(xv)))
> yv = predict(m1b, list(pH = phv, Biomass = xv), type = "response")
> lines(xv, yv, lty = 2)
> phv = factor(rep("high", length(xv)))
> yv = predict(m1b, list(pH = phv, Biomass = xv), type = "response")
> lines(xv, yv, lty = 4)
> legend("topright", with(species, levels(pH)), pch = PCH,
    title = "pH")
> mtext("log", line = -1.25)
> with(species, plot(Species ~ Biomass, pch = PCH[as.numeric(species$pH)]))
> phv = factor(rep("low", length(xv)))
> yv = predict(m2b, list(pH = phv, Biomass = xv), type = "response")
> lines(xv, yv, lty = 1)
> phv = factor(rep("mid", length(xv)))
> yv = predict(m2b, list(pH = phv, Biomass = xv), type = "response")
> lines(xv, yv, lty = 2)
> phv = factor(rep("high", length(xv)))
> yv = predict(m2b, list(pH = phv, Biomass = xv), type = "response")
> lines(xv, yv, lty = 4)
> legend("topright", with(species, levels(pH)), pch = PCH,
    title = "pH")
> mtext("identity", line = -1.25)
> with(species, plot(Species ~ Biomass, pch = PCH[as.numeric(species$pH)]))
> phv = factor(rep("low", length(xv)))
> yv = predict(m3b, list(pH = phv, Biomass = xv), type = "response")
> lines(xv, yv, lty = 1)
> phv = factor(rep("mid", length(xv)))
> yv = predict(m3b, list(pH = phv, Biomass = xv), type = "response")
> lines(xv, yv, lty = 2)
> phv = factor(rep("high", length(xv)))
> yv = predict(m3b, list(pH = phv, Biomass = xv), type = "response")
> lines(xv, yv, lty = 4)
> legend("topright", with(species, levels(pH)), pch = PCH,
    title = "pH")
> mtext("sqrt", line = -1.25)
```



Obrázek 12: Výsledky analýzy kovariance (různé průsečíky, stejná směrnice) pro různé linkovací funkce.

Nyní ověříme, zda se druhý model výrazně nezhoršil oproti předchozímu.

```
> anova(m1, m1b, test = "Chi")


Analysis of Deviance Table

Model 1: Species ~ Biomass * pH
Model 2: Species ~ Biomass + pH
  Resid. Df Resid. Dev Df Deviance P(>|Chi|)
1        84     83.201
2        86     99.242 -2  -16.040 0.0003288 ***
---
Signif. codes:  0 ,***, 0.001 ,**, 0.01 ,*, 0.05 ,., 0.1 , , 1
```

```
> anova(m2, m2b, test = "Chi")


Analysis of Deviance Table

Model 1: Species ~ Biomass * pH
Model 2: Species ~ Biomass + pH
  Resid. Df Resid. Dev Df Deviance P(>|Chi|)
1        84     67.070
2        86     68.126 -2  -1.0562    0.5897


> anova(m3, m3b, test = "Chi")


Analysis of Deviance Table

Model 1: Species ~ Biomass * pH
Model 2: Species ~ Biomass + pH
  Resid. Df Resid. Dev Df Deviance P(>|Chi|)
1        84     74.876
2        86     81.133 -2  -6.2566   0.04379 *
---
Signif. codes:  0 ,***, 0.001 ,**, 0.01 ,*, 0.05 ,., 0.1 , , 1
```

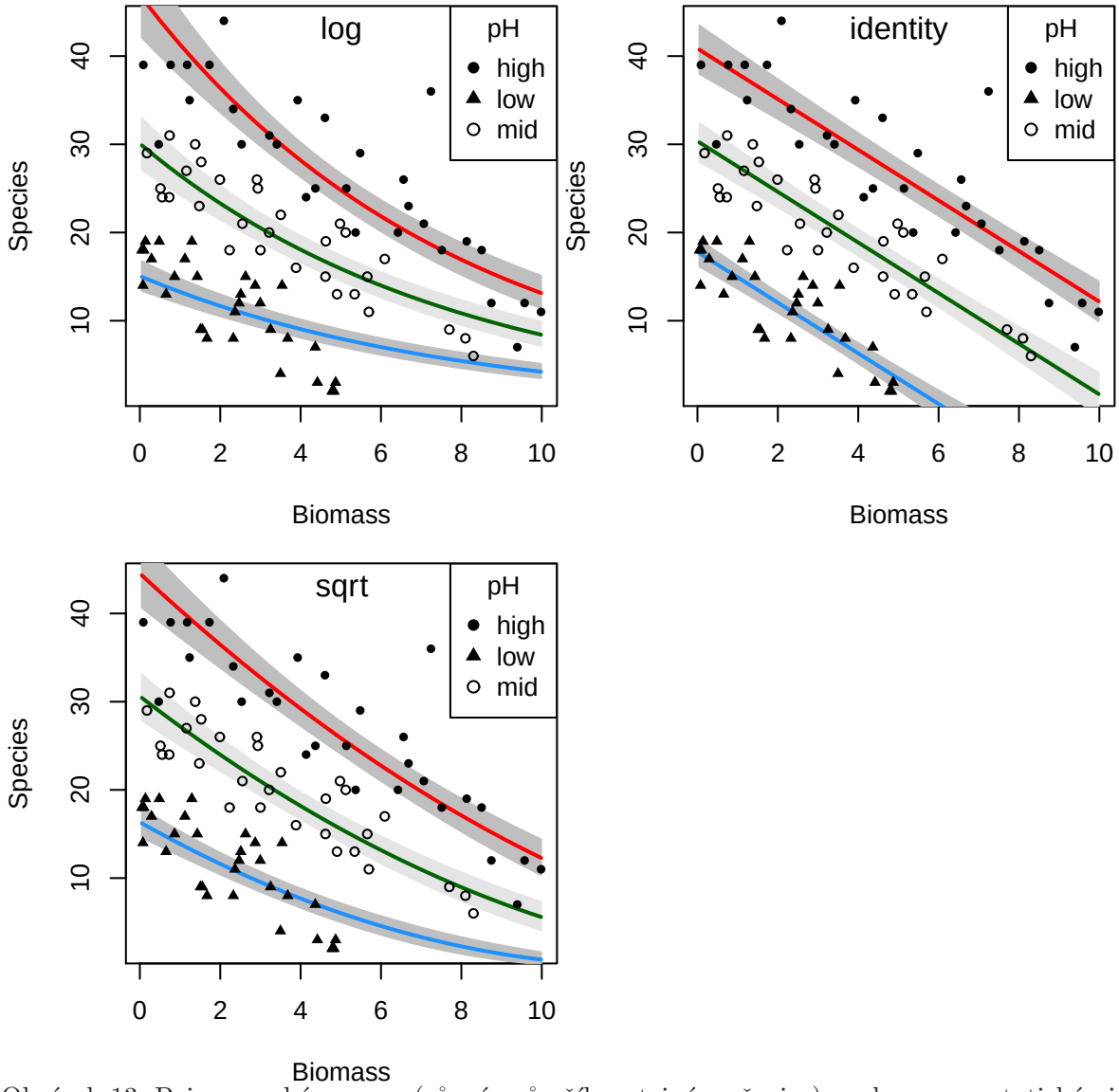
U modelů s logaritmickou a odmocninovou linkovací funkci výsledky testů ukazují, že vypuštěním jedné kovariáty se model významně zhoršil, pouze model s identickou link funkcí se zhoršení neprokázalo. Ale z výsledných grafů je zřejmé, že identická linkovací funkce není příliš vhodná.

Na závěr ještě výsledek vykreslíme také s intervaly spolehlivosti kolem středních hodnot.

```
> par(mfrow = c(2, 2), mar = c(4, 4, 1, 0) + 0.05)
> with(species, plot(Species ~ Biomass, type = "n"))
> ab <- with(species, range(Biomass))
> nn <- 200
> Levels <- with(species, levels(pH))
> Fac1 <- factor(rep(Levels[1], nn), levels = Levels)
> Fac2 <- factor(rep(Levels[2], nn), levels = Levels)
> Fac3 <- factor(rep(Levels[3], nn), levels = Levels)
> xx <- seq(ab[1], ab[2], length.out = nn)
> yy1 <- predict(m1b, list(Biomass = xx, pH = Fac1), type = "response")
> yy2 <- predict(m1b, list(Biomass = xx, pH = Fac2), type = "response")
> yy3 <- predict(m1b, list(Biomass = xx, pH = Fac3), type = "response")
> Predicted <- predict(m1b, list(Biomass = xx, pH = Fac1),
    type = "link", se = T)
> CI.L.log <- exp(Predicted$fit - 1.96 * Predicted$se.fit)
> CI.H.log <- exp(Predicted$fit + 1.96 * Predicted$se.fit)
> x <- c(xx, rev(xx))
> y <- c(CI.L.log, rev(CI.H.log))
> polygon(x, y, col = "gray75", border = "gray75")
> Predicted <- predict(m1b, list(Biomass = xx, pH = Fac2),
    type = "link", se = T)
> CI.L.log <- exp(Predicted$fit - 1.96 * Predicted$se.fit)
> CI.H.log <- exp(Predicted$fit + 1.96 * Predicted$se.fit)
```

```
> x <- c(xx, rev(xx))
> y <- c(CI.L.log, rev(CI.H.log))
> polygon(x, y, col = "gray75", border = "gray75")
> Predicted <- predict(m1b, list(Biomass = xx, pH = Fac3),
    type = "link", se = T)
> CI.L.log <- exp(Predicted$fit - 1.96 * Predicted$se.fit)
> CI.H.log <- exp(Predicted$fit + 1.96 * Predicted$se.fit)
> x <- c(xx, rev(xx))
> y <- c(CI.L.log, rev(CI.H.log))
> polygon(x, y, col = "gray90", border = "gray90")
> lines(xx, yy1, col = "red", lwd = 2)
> lines(xx, yy2, col = "dodgerblue", lwd = 2)
> lines(xx, yy3, col = "darkgreen", lwd = 2)
> with(species, points(Species ~ Biomass, cex = 0.75, pch = PCH[as.numeric(species$pH)]))
> legend("topright", with(species, levels(pH)), pch = PCH,
    title = "pH")
> mtext("log", line = -1.25)
> with(species, plot(Species ~ Biomass, type = "n"))
> yy1 <- predict(m2b, list(Biomass = xx, pH = Fac1), type = "response")
> yy2 <- predict(m2b, list(Biomass = xx, pH = Fac2), type = "response")
> yy3 <- predict(m2b, list(Biomass = xx, pH = Fac3), type = "response")
> Predicted <- predict(m2b, list(Biomass = xx, pH = Fac1),
    type = "link", se = T)
> CI.L.log <- (Predicted$fit - 1.96 * Predicted$se.fit)
> CI.H.log <- (Predicted$fit + 1.96 * Predicted$se.fit)
> x <- c(xx, rev(xx))
> y <- c(CI.L.log, rev(CI.H.log))
> polygon(x, y, col = "gray75", border = "gray75")
> Predicted <- predict(m2b, list(Biomass = xx, pH = Fac2),
    type = "link", se = T)
> CI.L.log <- (Predicted$fit - 1.96 * Predicted$se.fit)
> CI.H.log <- (Predicted$fit + 1.96 * Predicted$se.fit)
> x <- c(xx, rev(xx))
> y <- c(CI.L.log, rev(CI.H.log))
> polygon(x, y, col = "gray75", border = "gray75")
> Predicted <- predict(m2b, list(Biomass = xx, pH = Fac3),
    type = "link", se = T)
> CI.L.log <- (Predicted$fit - 1.96 * Predicted$se.fit)
> CI.H.log <- (Predicted$fit + 1.96 * Predicted$se.fit)
> x <- c(xx, rev(xx))
> y <- c(CI.L.log, rev(CI.H.log))
> polygon(x, y, col = "gray90", border = "gray90")
> lines(xx, yy1, col = "red", lwd = 2)
> lines(xx, yy2, col = "dodgerblue", lwd = 2)
> lines(xx, yy3, col = "darkgreen", lwd = 2)
> with(species, points(Species ~ Biomass, cex = 0.75, pch = PCH[as.numeric(species$pH)]))
> legend("topright", with(species, levels(pH)), pch = PCH,
    title = "pH")
> mtext("identity", line = -1.25)
> with(species, plot(Species ~ Biomass, type = "n"))
> yy1 <- predict(m3b, list(Biomass = xx, pH = Fac1), type = "response")
> yy2 <- predict(m3b, list(Biomass = xx, pH = Fac2), type = "response")
> yy3 <- predict(m3b, list(Biomass = xx, pH = Fac3), type = "response")
> Predicted <- predict(m3b, list(Biomass = xx, pH = Fac1),
    type = "link", se = T)
> CI.L.log <- (Predicted$fit - 1.96 * Predicted$se.fit)^2
> CI.H.log <- (Predicted$fit + 1.96 * Predicted$se.fit)^2
```

```
> x <- c(xx, rev(xx))
> y <- c(CI.L.log, rev(CI.H.log))
> polygon(x, y, col = "gray75", border = "gray75")
> Predicted <- predict(m3b, list(Biomass = xx, pH = Fac2),
    type = "link", se = T)
> CI.L.log <- (Predicted$fit - 1.96 * Predicted$se.fit)^2
> CI.H.log <- (Predicted$fit + 1.96 * Predicted$se.fit)^2
> x <- c(xx, rev(xx))
> y <- c(CI.L.log, rev(CI.H.log))
> polygon(x, y, col = "gray75", border = "gray75")
> Predicted <- predict(m3b, list(Biomass = xx, pH = Fac3),
    type = "link", se = T)
> CI.L.log <- (Predicted$fit - 1.96 * Predicted$se.fit)^2
> CI.H.log <- (Predicted$fit + 1.96 * Predicted$se.fit)^2
> x <- c(xx, rev(xx))
> y <- c(CI.L.log, rev(CI.H.log))
> polygon(x, y, col = "gray90", border = "gray90")
> lines(xx, yy1, col = "red", lwd = 2)
> lines(xx, yy2, col = "dodgerblue", lwd = 2)
> lines(xx, yy3, col = "darkgreen", lwd = 2)
> with(species, points(Species ~ Biomass, cex = 0.75, pch = PCH[as.numeric(species$pH)]))
> legend("topright", with(species, levels(pH)), pch = PCH,
    title = "pH")
> mtext("sqrt", line = -1.25)
```



Obrázek 13: Poissonovská regrese (různé průsečíky, stejná směrnice) spolu s asymptotickými intervaly spolehlivosti.