Homework9

Biologists were interested in understanding how the density of vegetation and height from the forest floor explain the number (count) of spiders found within 5 m radius plots in a Hawaiian wet forest. Height was measured as a categorical variable here. (see spiders dataset)

Import data

```
library(MASS)
library(car)
```

```
## Loading required package: carData
```

```
library(MuMIn)

ds <- readxl::read_excel('/Users/kanoalindiwe/Downloads/Projects/playground/R/Quantitative Ecology/Datas

# independent
vegdens <- ds$vegdens
height <- factor(ds$height)

# dependent
spiders <- ds$spiders
```

a) run the full model (with interaction) and check for overdispersion

```
# Initial model
model.test <- glm(spiders ~ vegdens + height + vegdens*height, family = "poisson")

# Outcome is a count
#Pass, spiders is count data

# Independence of observations
# Pass, assumed to be true

# Mean equals variance (equidispersion)
sum(resid(model.test, type = "pearson")^2)/model.test$df.resid
```

```
## [1] 2.436013
```

```
model.test$deviance/model.test$df.resid
```

```
## [1] 2.63312
```

```r
# Fail, value is above 1. Do negative binomial instead.

# Log of the mean is a linear function of predictors
# Skip to negative binomial

# Dispersion parameter
# Skip to negative binomial

# Zero-inflation or truncated
# Skip to negative binomial
```

b) if over-dispersed, use the negative binomial model to account for over-dispersion

```r
# Initial model
model.nb <- glm.nb(spiders ~ vegdens + height + vegdens*height)

# Count response (non-negative integers)
# Pass

# Independence of observations
# Pass, assumed

# Overdispersion present
mean(ds$spiders)
```

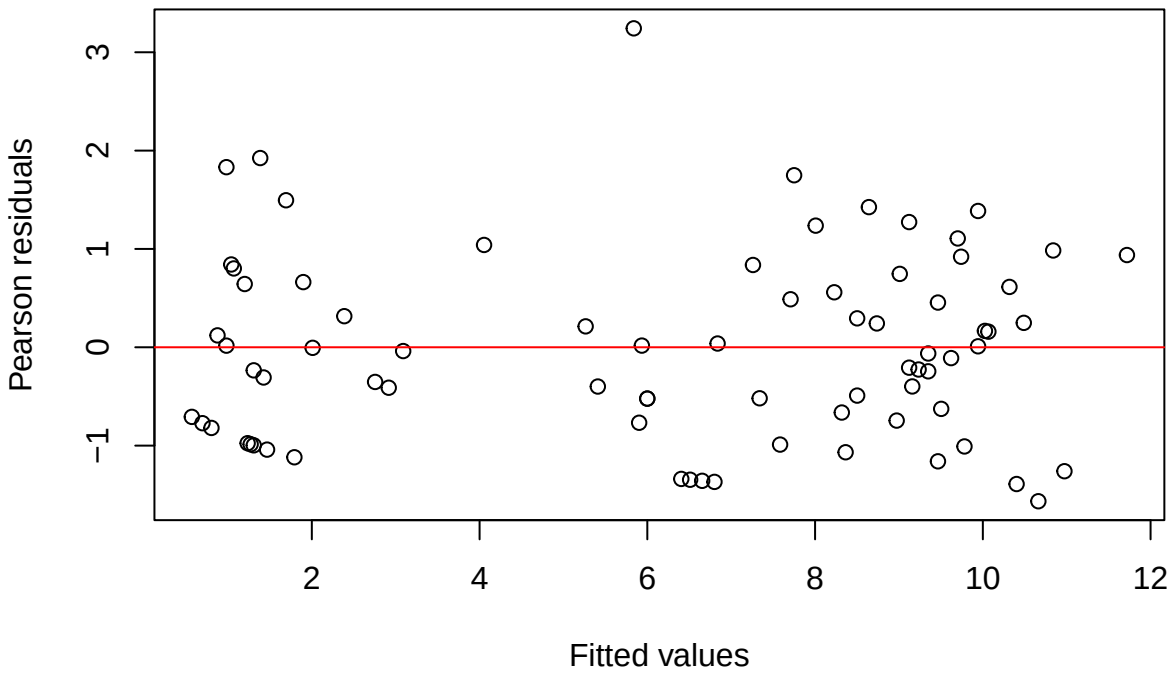
```
## [1] 6.25
```

```r
var(ds$spiders)
```

```
## [1] 29.44333
```

```r
# True, var is greater than mean

# Log link is appropriate
plot(fitted(model.nb), resid(model.nb, type = "pearson"),
     xlab = "Fitted values", ylab = "Pearson residuals")
abline(h = 0, col = "red")
```




```
# Pass, no curve or trend

# No severe multicollinearity among predictors
vif(model.nb)
```

```
## there are higher-order terms (interactions) in this model
## consider setting type = 'predictor'; see ?vif
```

```
##                     GVIF Df GVIF^(1/(2*Df))
## vegdens         26.03675  1        5.102622
## height          27.09487  2        2.281507
## vegdens:height 140.44697  2        3.442533
```

```
# Pass, values close to or under 5

# Zero-inflation or truncated
obs_zeros <- sum(spiders == 0)

pred <- predict(model.nb, type = "response")
exp_zeros <- sum(dnbinom(0, mu = pred, size = model.nb$theta))

obs_zeros
```

```
## [1] 8
```

```
exp_zeros
```

```
## [1] 8.214898
```

```
# Pass, observed and predicted are close
```

c) use model selection in package MuMIn to select best model. Be sure to assess importance values and report confidence intervals for the coefficients used in your final model

```
# Full model
model.nb.full <- glm.nb(spiders ~ vegdens + height + vegdens*height, na.action = "na.fail")

# Dredge
model.set <- dredge(model.nb.full)
```

```
## Fixed term is "(Intercept)"
```

```
summary(model.set)
```

```
##   (Intercept)      height     vegdens          height:vegdens       df
##  Min.   :0.4055   +   :3   Min.   :-0.028657   +   :1         Min.   :2.0
##  1st Qu.:0.5712   NA's:2   1st Qu.:-0.016741   NA's:4         1st Qu.:3.0
##  Median :1.3571            Median :-0.004825                  Median :4.0
##  Mean   :1.2185            Mean   :-0.011845                  Mean   :4.2
##  3rd Qu.:1.8326            3rd Qu.:-0.003439                  3rd Qu.:5.0
##  Max.   :1.9260            Max.   :-0.002053                  Max.   :7.0
##                            NA's   :2
##      logLik            AICc           delta              weight
##  Min.   :-220.1   Min.   :395.9   Min.   : 0.0000   Min.   :0.0000
##  1st Qu.:-220.0   1st Qu.:396.0   1st Qu.: 0.1517   1st Qu.:0.0000
##  Median :-193.8   Median :396.2   Median : 0.3200   Median :0.3066
##  Mean   :-203.3   Mean   :415.8   Mean   :19.8743   Mean   :0.2000
##  3rd Qu.:-192.6   3rd Qu.:444.4   3rd Qu.:48.5249   3rd Qu.:0.3336
##  Max.   :-190.1   Max.   :446.3   Max.   :50.3752   Max.   :0.3598
##
```

```
model.set
```

```
## Global model call: glm.nb(formula = spiders ~ vegdens + height + vegdens * height,
##     na.action = "na.fail", init.theta = 4.145679499, link = log)
## ---
## Model selection table
##    (Int) hgh       vgd hgh:vgd df   logLik  AICc delta weight
## 8 1.3570   + -0.028660       +  7 -190.117 395.9  0.00  0.360
## 4 0.5712   + -0.004825          5 -192.588 396.0  0.15  0.334
## 2 0.4055   +                     4 -193.819 396.2  0.32  0.307
## 1 1.8330                         2 -220.121 444.4 48.52  0.000
## 3 1.9260     -0.002053           3 -219.961 446.3 50.38  0.000
## Models ranked by AICc(x)
```

```r
# Three models should be averaged

# Select models
model.avg.out <- model.avg(model.set, subset = delta < 2)
summary(model.avg.out)
```

```
##
## Call:
## model.avg(object = model.set, subset = delta < 2)
##
## Component model call:
## glm.nb(formula = spiders ~ <3 unique rhs>, na.action = na.fail,
##      init.theta = 4.145679499, link = log)
##
## Component models:
##     df  logLik   AICc delta weight
## 123  7 -190.12 395.88  0.00   0.36
## 12   5 -192.59 396.03  0.15   0.33
## 1    4 -193.82 396.20  0.32   0.31
##
## Term codes:
##         height        vegdens height:vegdens
##              1              2              3
##
## Model-averaged coefficients:
## (full average)
##                   Estimate Std. Error Adjusted SE z value Pr(>|z|)
## (Intercept)       0.803174   0.531472    0.534952   1.501  0.13325
## heightL           1.386059   0.499995    0.504992   2.745  0.00606 **
## heightM           1.489107   0.635323    0.639010   2.330  0.01979 *
## vegdens          -0.011921   0.015062    0.015139   0.787  0.43102
## heightL:vegdens   0.008352   0.013827    0.013913   0.600  0.54830
## heightM:vegdens   0.011789   0.018088    0.018167   0.649  0.51636
##
## (conditional average)
##                  Estimate Std. Error Adjusted SE z value Pr(>|z|)
## (Intercept)       0.80317    0.53147     0.53495   1.501  0.13325
## heightL           1.38606    0.49999     0.50499   2.745  0.00606 **
## heightM           1.48911    0.63532     0.63901   2.330  0.01979 *
## vegdens          -0.01719    0.01538     0.01549   1.110  0.26699
## heightL:vegdens   0.02321    0.01365     0.01389   1.671  0.09480 .
## heightM:vegdens   0.03276    0.01490     0.01516   2.161  0.03072 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Spider abundance varies significantly with height above the forest floor, with more spiders found in the Low and Mid height strata compared to the reference height. Vegetation density on its own does not strongly influence spider counts.