P8131-hw5-rw2844

Renjie Wei

2022-03-22

Question 1

(a) Fit a Poisson model (M1) with log link with W as the single predictor. Check the goodness of fit and interpret your model.

Model 1 (M1):

$$\log(Sa_i) = \log(1) + \beta_0 + \beta_1 \times \mathrm{W}_i$$

Given the value of the residual deviance statistic of 567.8785725 with $df = 171$, the p-value is 0, so the model does not fit well. However, since the Model 1 shows that $\beta_1 = 0.164$, which means the wider the female crab, the greater expected number of male satellites. More specifically, for one unit of increase in the width (W), the number of Satellites (Sa) will increase by 1.178, and the 95% confidence interval (CI) is (1.133 ,1.225) .

(b) Fit a model (M2) with W and Wt as predictors. Compare it with the model in (a). Interpret your results.

Model 2 (M2):

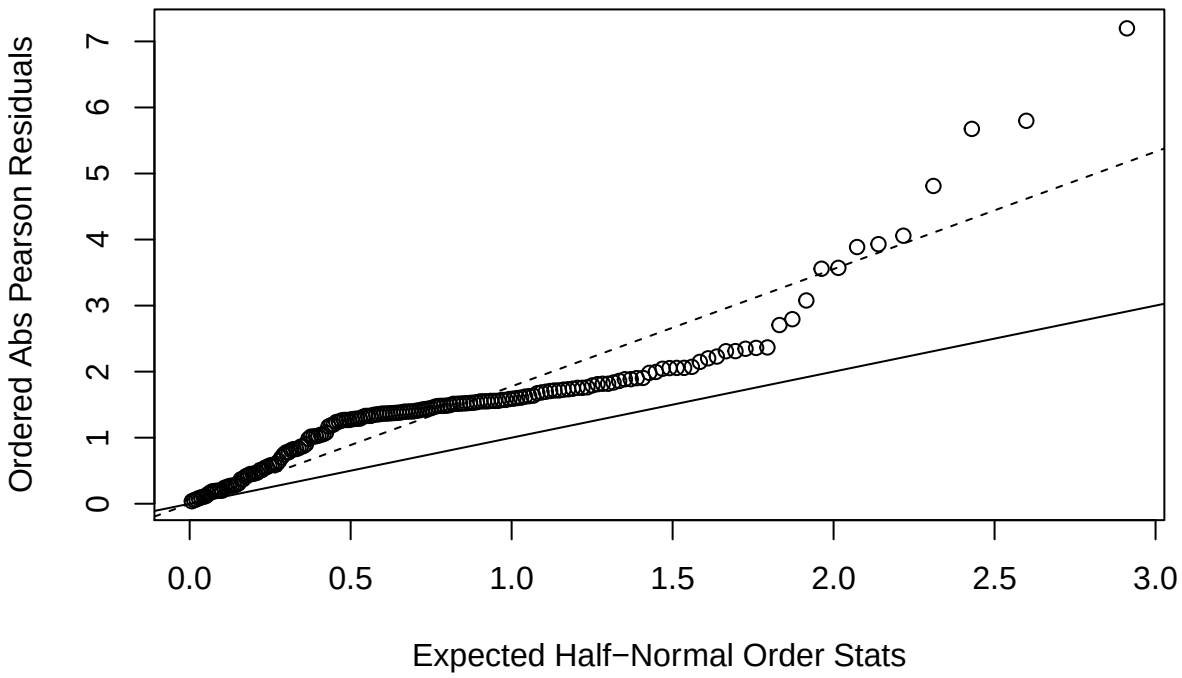
$$\log(Sa_i) = \log(1) + \beta_0 + \beta_1 \times \mathrm{W}_i + \beta_2 \times \mathrm{Wt}_i$$

The M1 and M2 are nested models, comparing the two model is equivalent to testing $H_0 : \beta_2 = 0$ vs. $H_1 : \beta_2 \neq 0$. The test statistic is 7.993392 with $df = 1$, the p-value is 0.0047<0.05. Null hypothesis is rejected, which means M2 better fits the data. However, given the value of the residual deviance statistic of 559.8851805 with $df = 170$, the p-value is 0, so the model does not fit well.

The Model 2 shows that $\beta_1 = 0.046$ and $\beta_2 = 0.447$, which means the wider and heavier the female crab, the greater expected number of male satellites. More specifically, holding other conditions fixed, for one unit of increase in the width (W), the number of Satellites (Sa) will increase by 1.047, and the 95% confidence interval (CI) is (0.955 ,1.147); Holding other conditions fixed, for one unit of increase in the weight (Wt), the number of Satellites (Sa) will increase by 1.564, and the 95% confidence interval (CI) is (1.147 ,2.135) .

(c) Check over dispersion in M2. Interpret the model after adjusting for over dispersion.

Deviance and Pearson χ^2 statistics of M2 are large and from (b) we know M2 is lack of fit.



The half-normal plot using residual from this model shows evidence of over-dispersion.

The dispersion parameter ϕ is 3.156 and $\tilde{\phi}$ is 3.293, both are greater than 1.

```
##
## Call:
## glm(formula = Sa ~ W + Wt, family = poisson, data = crab.dat)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.9308  -1.9705  -0.5481   0.9700   4.9905
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.29168    1.59771  -0.808    0.419
## W            0.04590    0.08309   0.552    0.581
## Wt           0.44744    0.28184   1.588    0.112
##
## (Dispersion parameter for poisson family taken to be 3.156449)
##
##     Null deviance: 632.79  on 172  degrees of freedom
## Residual deviance: 559.89  on 170  degrees of freedom
## AIC: 921.18
##
## Number of Fisher Scoring iterations: 6
```

The model fitted with constant over-dispersion parameter ϕ, the deviance analysis shows a p-value 0.333, which means this model fit the data well.

Question 2

(a) Fit a Poisson model with log link to the data with area, year, and length as predictors. Interpret each model parameter.

Model:

$$\log(\text{Intensity}_i) = \log(1)+\beta_0+\beta_1\times\text{Area2}+\beta_2\times\text{Area3}+\beta_3\times\text{Area4}+\beta_4\times\text{Year2000}+\beta_5\times\text{Year2001}+\beta_6\times\text{Length}$$

The Model shows that:

For parameters about **Area**:

β_1 = -0.212. Holding other conditions fixed, Area change from 1 to 2, the Intensity decreased by 0.809, and the 95% confidence interval (CI) is (0.734 ,0.891).

β_2 = -0.117. Holding other conditions fixed, Area change from 1 to 3, the Intensity decreased by 0.89, and the 95% confidence interval (CI) is (0.818 ,0.968).

β_3 = 1.405. Holding other conditions fixed, Area change from 1 to 3, the Intensity increased by 4.075, and the 95% confidence interval (CI) is (3.802 ,4.372).

For parameters about **Year**:

β_4 = 0.67. Holding other conditions fixed, Year change from 1999 to 2000, the Intensity increased by 1.955, and the 95% confidence interval (CI) is (1.85 ,2.065).

β_5 = -0.218. Holding other conditions fixed, Year change from 1999 to 2001, the Intensity decreased by 0.804, and the 95% confidence interval (CI) is (0.76 ,0.851).

For the parameter about **Length**:

β_6 = -0.028. Holding other conditions fixed, Year change from 1999 to 2001, the Intensity decreased by 0.972, and the 95% confidence interval (CI) is (0.97 ,0.974).

(b) Test for goodness of fit of the model in (a) and state conclusions.

Given the value of the residual deviance statistic of 1.9152798×10^4 with $df = 1184$, the p-value is 0, so the model does not fit well.

(c) Researchers suspect that there may be two strains of fish, one that is susceptible to parasites and one that is not. Without knowing which fish are susceptible, this could be regarded as a zero-inflated model. Building on the model in (a) (using the same predictors), fit an appropriate model to the data that can account for extra zeros. Provide an interpretation for each model parameter in terms of the problem.

In this problem, I assume that `Area` is the factor that determine the strain of fishZ_i, where $Z_i = 0$ with $P(Z_i = 0) = \pi_i$means the strain of fish is not susceptible to parasite. And the distribution of Intensity conditioned on strain of fish is $Y_i|(Z_i = 0) = 0,\ Y_i|(Z_i = 1) = Pois(\lambda_i)$ So the models are:

$$\log(\frac{\pi_i}{1-\pi_i}) = \beta_{01} + \beta_{11} \times \text{Area2} + \beta_{21} \times \text{Area3} + \beta_{31} \times \text{Area4}$$

and

$$\log(\text{Intensity}_i) = \beta_{02} + \beta_{12} \times \text{Year2000} + \beta_{22} \times \text{Year2001} + \beta_{32} \times \text{Length}$$

```
##
## Call:
## zeroinfl(formula = Intensity ~ Year + Length | Area, data = para.dat)
##
## Pearson residuals:
##     Min      1Q  Median      3Q     Max
## -1.5077 -0.7131 -0.6447 -0.2369 26.2175
##
## Count model coefficients (poisson with log link):
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)  4.6630522  0.0459573 101.465  < 2e-16 ***
## Year2000     0.4214742  0.0278972  15.108  < 2e-16 ***
## Year2001     0.0988373  0.0286162   3.454 0.000553 ***
## Length      -0.0438777  0.0009298 -47.193  < 2e-16 ***
##
## Zero-inflation model coefficients (binomial with logit link):
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.001797   0.121809   0.015    0.988
## Area2        0.746780   0.183065   4.079 4.52e-05 ***
## Area3        0.680875   0.161795   4.208 2.57e-05 ***
## Area4       -0.882654   0.180987  -4.877 1.08e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Number of iterations in BFGS optimization: 11
## Log-likelihood: -7563 on 8 Df
```

For the logistic model of the binary latent variable Strains Z_i:

β_{12} = 0.747. Holding other conditions fixed, Area change from 1 to 2, the odds of the fish belongs to the parasite insusceptible strain increased by 2.11, and the 95% confidence interval (CI) is (1.474 ,3.021).

β_{22} = 0.681. Holding other conditions fixed, Area change from 1 to 3, the odds of the fish belongs to the parasite insusceptible strain increased by 1.976, and the 95% confidence interval (CI) is (1.439 ,2.713).

β_{32} = -0.883. Holding other conditions fixed, Area change from 1 to 4, the odds of the fish belongs to the parasite insusceptible strain decreased by 0.414, and the 95% confidence interval (CI) is (0.29 ,0.59).

For the poisson model of the poisson response Intensity $Y_i|(Z_i = 1) = Pois(\lambda_i)$:

β_{11} = 0.421. Holding other conditions fixed, Year change from 1999 to 2000, the Intensity increased by 1.524, and the 95% confidence interval (CI) is (1.443 ,1.61).

β_{21} = 0.099. Holding other conditions fixed, Year change from 1999 to 2001, the Intensity increased by 1.104, and the 95% confidence interval (CI) is (1.044 ,1.168).

β_{31} = -0.044. Holding other conditions fixed, for each unit increase in Length, the Intensity decreased by 4.075, and the 95% confidence interval (CI) is (0.955 ,0.959).