**Argrett Lab 6**

For this lab, we will continue working with the potato virus Z data, testing the hypothesis that disease prevalence is driven by aphid abundance using a similar hurdle model (36 pts total):

1. We need to standardize aphid abundance measured at each farm by the trapping effort. Make a new variable that is the number of aphids divided by the number of traps. Turn in just the code for this. (2 pts)

PVY$APT <- PVY$Naphids/PVY$NTraps

1. Construct the first step of the hurdle model to test the hypothesis of whether the presence/absence of infection depends on the effects of Year, Standardized Aphid Abundance, and their interaction. Turn in just the code to state this model. (2pts)

model <- glm(inf~Year\*APT, data=PVY, family="binomial")

* 1. Chart, scatter chart

     Description automatically generatedChart, scatter chart

     Description automatically generatedCheck model assumptions by making a qqplot with residuals from qresid(), making a residual~fitted plot to check for dispersion patterns, and calculate ss:df ratio to check for overdispersion (6 pts). Give your best judgement/interpretation of the output. Are model assumptions upheld? (2pts)

Model assumptions can be upheld due to the fact that residuals are relatively evenly distributed above and below the fitted line. Calculating for overdispersion, we see an indication of slight overdispersion (ss:df ratio = 1.047588) but not enough to reject model assumptions.

* 1. Interpret the output of the presence-absence model using summary() and anova(). How would you report these results? (2 pts for the output, 4 pts for results sentences)

**summary(model)**

Call:

glm(formula = inf ~ Year \* APT, family = "binomial", data = PVY)

Deviance Residuals:

Min 1Q Median 3Q

-1.6259 -1.1543 0.7047 0.9403

Max

1.4485

Coefficients:

Estimate Std. Error z value

(Intercept) -1.0389 0.9375 -1.108

Year2 1.1307 1.2684 0.891

APT 1.3971 0.9112 1.533

Year2:APT -0.7349 1.2460 -0.590

Pr(>|z|)

(Intercept) 0.268

Year2 0.373

APT 0.125

Year2:APT 0.555

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 52.925 on 39 degrees of freedom

Residual deviance: 48.265 on 36 degrees of freedom

AIC: 56.265

Number of Fisher Scoring iterations: 4

**Anova(model, test = “Chisq”)**

Analysis of Deviance Table

Model: binomial, link: logit

Response: inf

Terms added sequentially (first to last)

Df Deviance Resid. Df

NULL 39

Year 1 0.3275 38

APT 1 3.9917 37

Year:APT 1 0.3411 36

Resid. Dev Pr(>Chi)

NULL 52.925

Year 52.598 0.56711

APT 48.606 0.04573 \*

Year:APT 48.265 0.55920

---

Signif. codes:

0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’

0.1 ‘ ’ 1

Looking at the results of our two-step hurdle model, we found that probability of infection presence did not depend on year, or aphid abundance by year (both P>0.05, pseudo - R2 = 0.088). However there was a significant relationship between aphid abundance and probability of infection presence (X2 1 and 37 df, P = 0.0457).

1. Construct the second step of the hurdle model using the subset of data that only has farms with infected plants. Test the hypothesis that disease prevalence depends on the effects of Year, Standardized Aphid Abundance, and their interaction. Turn in just the code to state this model. (2pts)

PVY$Infect<-ifelse(PVY$Prev>0,1,0)

prevPVY <- subset(PVY, Infect>0)

model2<-glm(Infect~Year\*APT, data=prevPVY, family="binomial")

* 1. Chart, scatter chart

     Description automatically generatedCheck model assumptions by making a qqplot with residuals from qresid(), making a residual~fitted plot to check for dispersion patterns, and calculate ss:df ratio to check for overdispersion (6 pts). Give your best judgement/interpretation of the output. Are model assumptions upheld? (2pts)

Chart, scatter chart

Description automatically generatedResiduals seem evenly distributed above and below the line however our it is over dispersed 4.189722. For now, we will continue onwards.

* 1. Interpret the output of the prevalence model using summary() and anova(). How would you report these results? You may find it easier to come back to the results writing after making figures to better visualize the interaction (2 pts for the output, 4 pts for results sentences)

**summary(model2)**

Call:

glm(formula = Prev ~ Year \* APT, family = "binomial", data = prevPVY,

weights = NPlants)

Deviance Residuals:

Min 1Q Median 3Q Max

-3.662 -1.055 -0.495 1.241 4.989

Coefficients:

Estimate Std. Error z value Pr(>|z|)

(Intercept) -2.7772 0.3290 -8.442 < 2e-16 \*\*\*

Year2 1.9990 0.4214 4.744 2.09e-06 \*\*\*

APT 0.6700 0.1881 3.563 0.000367 \*\*\*

Year2:APT -1.8672 0.3599 -5.189 2.12e-07 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 113.728 on 24 degrees of freedom

Residual deviance: 79.905 on 21 degrees of freedom

AIC: 159.24

Number of Fisher Scoring iterations: 5

**Anova(model2, type = “Caisq”)**

Analysis of Deviance Table

Model: binomial, link: logit

Response: Prev

Terms added sequentially (first to last)

Df Deviance Resid. Df Resid. Dev Pr(>Chi)

NULL 24 113.728

Year 1 0.052 23 113.677 0.8202

APT 1 0.383 22 113.293 0.5357

Year:APT 1 33.388 21 79.905 7.548e-09 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Looking at the results of our two-step hurdle model we also found that year and aphid abundance had no significant relationship with the prevalence of PVY infection (both P>0.05, pseudo-R2 = 0.297). There was a significant relationship between the interaction of year and aphid abundance on PVY infection prevalence (X2 on 1 and 21 df, P = 7.548e-09)

1. Chart, diagram

   Description automatically generatedMake a publication-quality two-panel plot similar to that in the example code, with panel A showing the raw data and fitted model for the presence-absence model, and panel B showing the raw data and fitted model for the prevalence model. Include a figure caption (4 pts).

Figure 1: Data and fitted hurdle models for (A) probability of infection   
presence and (B) prevalence of infection given its presence depending on year and standardized aphid abundance. Model statistics are given in panel. Points show raw data, and line and ribbons show fitted model with SE.