04\_model-analysis\_infection-endpoint\_status

Bryony Allen

14/05/2019

# Part 1: import data

**N.B.** At the moment this code chuck reads in a .csv file with endpoint infection data, originally compiled in excel and then cleaned in “02\_tidy\_data-qpcr”script. In the future you will read in the .csv file of merged qPCR outputs (created, checked and cleaned in “02\_tidy\_data-qpcr”script) and experiment metadata (checked and cleaned in “02\_tidy\_data-metadata”script).

## Observations: 321  
## Variables: 26  
## $ ID <fct> A1.1, A1.2, A1.3, A1.4, A1.5, A1.6, A1...  
## $ Species <fct> Bb, Bb, Bb, Bb, Bb, Bb, Bb, Bb, Bb, Bb...  
## $ ExperimentNo <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,...  
## $ Scenario <fct> Coinfection, Coinfection, Coinfection,...  
## $ Treatment <fct> Rv-Bd, Rv-Bd, Rv-Bd, Rv-Bd, Rv-Bd, Rv-...  
## $ Exposure.1 <fct> rv, rv, rv, rv, rv, rv, rv, rv, rv, rv...  
## $ Exposure.2 <fct> bd, bd, bd, bd, bd, bd, bd, bd, bd, bd...  
## $ endpoint.date <fct> 06/06/2018, 26/05/2018, 06/06/2018, 27...  
## $ endpoint.code <fct> EU, MORT, EU, MORT, EU, MORT, EU, MORT...  
## $ Bd.endpoint.status <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,...  
## $ Bd.endpoint.CTmean <dbl> 30.94989, 37.30877, 32.19301, 36.65141...  
## $ Bd.endpoint.Qmean <dbl> 23.6021347, 0.4950310, 11.0839987, 0.7...  
## $ Rv.MCPendpoint.status <int> 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 0, 0,...  
## $ Rv.MCPendpoint.CTmean <dbl> 0.00000, 24.75441, 0.00000, 22.90014, ...  
## $ Rv.MCPendpoint.Qmean <dbl> 0.000000e+00, 2.462007e+05, 0.000000e+...  
## $ Rv.EBF3Nendpoint.status <int> 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 0, 0,...  
## $ Rv.EBF3Nendpoint.CTmean <dbl> NA, 33.46544, NA, 31.07020, NA, 34.558...  
## $ Rv.EBF3Nendpoint.Qmean <dbl> NA, 357.7631, NA, 1496.9165, NA, 182.4...  
## $ EMA.YN <int> 1, 0, 1, 0, 1, 1, 1, NA, NA, NA, NA, N...  
## $ EMA.date <fct> , NA, , NA, , , , NA, NA, NA, NA, NA, ...  
## $ EMA.GE.EMA <dbl> 0.004041085, NA, 0.479822159, NA, 0.00...  
## $ EMA.GE.WS <dbl> 0.026650012, NA, 0.005265172, NA, 0.00...  
## $ Bd.endpoint.GE <dbl> 236.021347, 4.950310, 110.839987, 7.37...  
## $ viable.GE <dbl> 0.004041085, NA, 0.479822159, NA, 0.00...  
## $ dead.GE <dbl> 0.022608927, NA, -0.474556987, NA, -0....  
## $ Rv.endpoint.load <dbl> NA, 1.376334e+03, NA, 9.135792e+02, NA...

# Part 2: visualise data

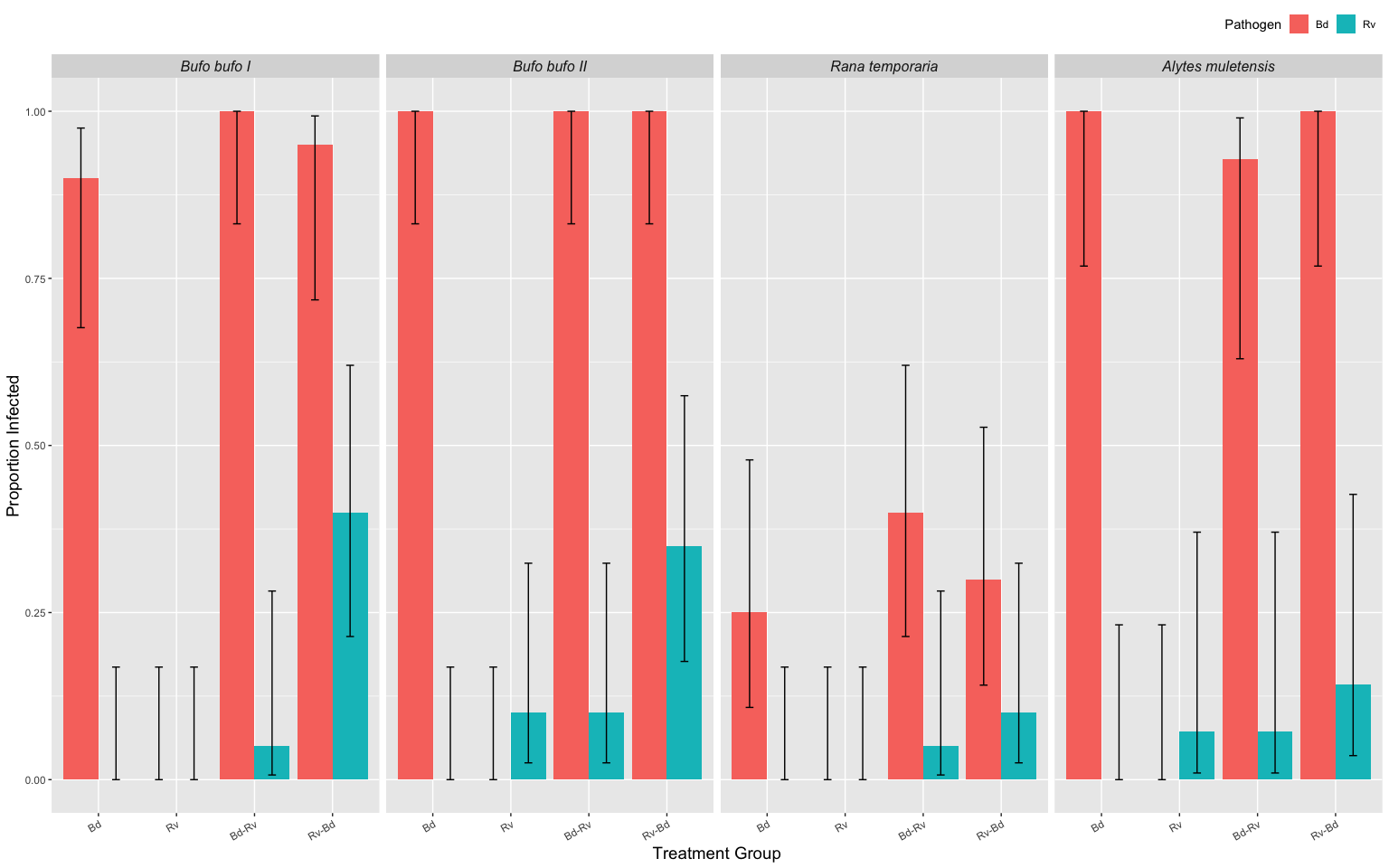


Fig.1. Proportion of individuals infected, by pathogen, within a treatment group for the three host species. Please note that Alytes II has been excluded.

**N.B.** *Bufo bufo* II have 2 individuals with Rv infection status in the Rv only group. These two records are sketchy as the qPCR results were inconclusive first time and then low the second time. Interestingly, the *Alytes muletensis* have the same pattern of a few individuals infected with ranavirus in the Rv only group with an equivalent number in the Bd-Rv treatment group. ?? susceptibility pattern or timing of Rv dose pattern ???

# Part 3: Bd: apply models to endpoint infection status

Binomial GLM’s where  
 response variable = Endpoint status [binary; 0,1]  
 explanatory variable(s) = Treatment [categorical] & ExperimentNo. [categorical]

**N.B.** I use ExperimentNo as a proxy for species where

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Experiment No. | Species | total Bd zsp’s | min. temp. (oC) | max. temp. (oC) |
| 1 | *Bufo bufo* I | 3,675,000 | 16.6 | 23.5 |
| 2 | *Bufo bufo* II | 1,443,750 | 16.7 | 27.6 |
| 3 | *Rana temporaria* | 2,336,250 | 16.7 | 27.6 |
| 4 | *Alytes muletensis* I | 472,500 | 15 | 16.6 |
| 5 | *Alytes muletensis* II | 294,759 | 15 | 16.6 |

… as this also accounts for Bd dose and room temperature variation between experiments.

## Part 3a: Endpoint Infection Status: Bd

Here I create a dataframe with the Rv-only treatment group removed (as they have never been exposed to Bd) and without *Alytes muletensis* babies as they only have one treatment group which I analyse seperately later.

Bd.status <- data.endpoint %>%  
 filter(!Treatment=="Rv") %>% # removal of Rv only treatment group   
 filter(!ExperimentNo=='5') %>% # removal of Alytes babies  
 mutate(ExperimentNo = as.factor(ExperimentNo)) %>%  
 dplyr::select(ID, Species, ExperimentNo, Treatment, Bd.endpoint.status, Bd.endpoint.GE)   
  
unique(data.endpoint$Treatment)  
droplevels(Bd.status)

## Part 3a: Bd: model comparison

The four models:

Bd.status1 <- glm(Bd.endpoint.status ~ Treatment \* ExperimentNo, data=Bd.status, family=binomial)  
  
Bd.status2 <- glm(Bd.endpoint.status ~ Treatment + ExperimentNo, data=Bd.status, family=binomial)  
  
Bd.status3 <- glm(Bd.endpoint.status ~ Treatment, data=Bd.status, family=binomial)  
  
Bd.status4 <- glm(Bd.endpoint.status ~ ExperimentNo, data=Bd.status, family=binomial)  
  
Bd.status.N <- glm(Bd.endpoint.status ~ 1, data=Bd.status, family=binomial)

Models compared using analysis of deviance with test='Chi' selected because of the binomial error family.

anova(Bd.status1, Bd.status2, test="Chisq") # start by comparing the interaction terms

## Analysis of Deviance Table  
##   
## Model 1: Bd.endpoint.status ~ Treatment \* ExperimentNo  
## Model 2: Bd.endpoint.status ~ Treatment + ExperimentNo  
## Resid. Df Resid. Dev Df Deviance Pr(>Chi)  
## 1 210 102.00   
## 2 216 106.75 -6 -4.7578 0.5752

anova(Bd.status2, Bd.status3, test="Chisq") # compares Trt and Species to just Trt

## Analysis of Deviance Table  
##   
## Model 1: Bd.endpoint.status ~ Treatment + ExperimentNo  
## Model 2: Bd.endpoint.status ~ Treatment  
## Resid. Df Resid. Dev Df Deviance Pr(>Chi)   
## 1 216 106.75   
## 2 219 223.16 -3 -116.41 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

anova(Bd.status2, Bd.status4, test="Chisq") # compares Trt and Species to just Species

## Analysis of Deviance Table  
##   
## Model 1: Bd.endpoint.status ~ Treatment + ExperimentNo  
## Model 2: Bd.endpoint.status ~ ExperimentNo  
## Resid. Df Resid. Dev Df Deviance Pr(>Chi)  
## 1 216 106.75   
## 2 218 108.19 -2 -1.4378 0.4873

anova(Bd.status4, Bd.status.N, test="Chisq") # compares Species model to null model

## Analysis of Deviance Table  
##   
## Model 1: Bd.endpoint.status ~ ExperimentNo  
## Model 2: Bd.endpoint.status ~ 1  
## Resid. Df Resid. Dev Df Deviance Pr(>Chi)   
## 1 218 108.19   
## 2 221 223.83 -3 -115.64 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

ANOVA 1: suggests we should reject the more complex model with interaction terms (Treatment \* ExperimentNo) in favour for the model with just the terms (pvalue = 0.5752)

ANOVA 2: suggests we should favour the more complex model (Treatment + ExperimentNo) over the model with Treatment only (pvalue = < .001) so ExperimentNo leads to significantly improved fit

ANOVA 3: suggests we should reject the more complex model (Treatment + ExperimentNo) and choose the model with just ExperimentNo. (pvalue = 0.4873)

ANOVA 4: suggests we should favour the more complex model (ExperimentNo.) over the null model (pvalue = < .001) so ExperimentNo leads to significantly improved fit

**Conclusion**: we should choose the model with just ExperimentNo (aka species)

## Part 3b: Bd: model fit

## ID Species ExperimentNo Treatment Bd.endpoint.status Bd.endpoint.GE  
## 1 A1.1 Bb 1 Rv-Bd 1 236.021347  
## 2 A1.2 Bb 1 Rv-Bd 1 4.950310  
## 3 A1.3 Bb 1 Rv-Bd 1 110.839987  
## 4 A1.4 Bb 1 Rv-Bd 1 7.370029  
## 5 A1.5 Bb 1 Rv-Bd 1 189.975014  
## 6 A1.6 Bb 1 Rv-Bd 1 134.006043  
## bi.glm  
## 1 0.95  
## 2 0.95  
## 3 0.95  
## 4 0.95  
## 5 0.95  
## 6 0.95

## Part 3c: Bd: model plotting

Bd.status4 <- glm(Bd.endpoint.status ~ ExperimentNo, data=Bd.status, family=binomial)  
  
# create a dataframe of "new" data   
newdat <- expand.grid(ExperimentNo=c("1", "2", "3", "4"),Treatment=c("Bd", "Bd-Rv", "Rv-Bd"))  
  
# predict the value/result of the new data using the glm  
newdat <-cbind(newdat, predict(object = Bd.status4, # the model   
 newdata=newdat, se=TRUE, type="response", print.matrix=T)) # dataframe of new data   
newdat

## ExperimentNo Treatment fit se.fit residual.scale  
## 1 1 Bd 0.9500000 2.813657e-02 1  
## 2 2 Bd 1.0000000 4.416308e-06 1  
## 3 3 Bd 0.3166667 6.005399e-02 1  
## 4 4 Bd 0.9761905 2.352437e-02 1  
## 5 1 Bd-Rv 0.9500000 2.813657e-02 1  
## 6 2 Bd-Rv 1.0000000 4.416308e-06 1  
## 7 3 Bd-Rv 0.3166667 6.005399e-02 1  
## 8 4 Bd-Rv 0.9761905 2.352437e-02 1  
## 9 1 Rv-Bd 0.9500000 2.813657e-02 1  
## 10 2 Rv-Bd 1.0000000 4.416308e-06 1  
## 11 3 Rv-Bd 0.3166667 6.005399e-02 1  
## 12 4 Rv-Bd 0.9761905 2.352437e-02 1

expl.var <- c(1:3) # chose the range for the x-axis (Treatment)  
exp.labs <- c("1" = "Bufo bufo I", "2" = "Bufo bufo II", "3" = "Rana temporaria", "4" = "Alytes muletensis")  
  
# subset the data so you can plot each seperatly   
newdat1<- subset(newdat, ExperimentNo== "1")   
newdat2<- subset(newdat, ExperimentNo=="2")  
newdat3<- subset(newdat, ExperimentNo=="3")  
newdat4<- subset(newdat, ExperimentNo=="4")

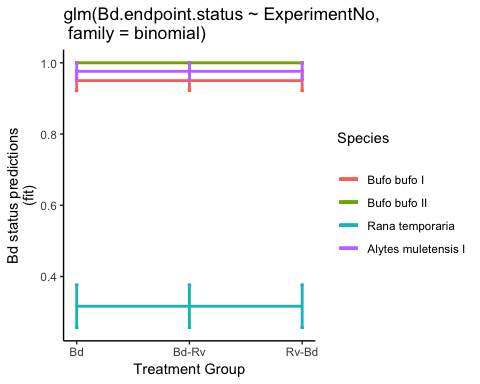
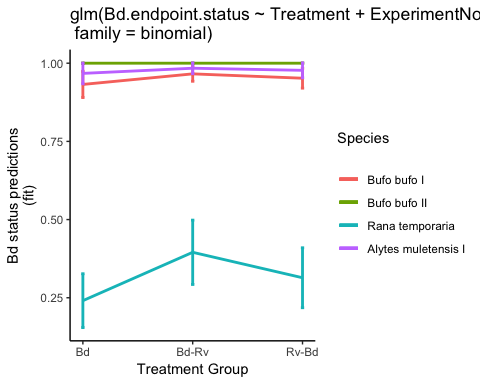


Fig. 2. Probability of Bd infection status (GE) predicted by model

**To Do** play with plots - check plot with Experiment No. as explanatory variable

This is the model predictions for Bd status with ExperimentNo. and Treatment

## ExperimentNo Treatment fit se.fit residual.scale  
## 1 1 Bd 0.9321182 4.194738e-02 1  
## 2 2 Bd 1.0000000 6.122604e-06 1  
## 3 3 Bd 0.2406850 8.609302e-02 1  
## 4 4 Bd 0.9674240 3.322510e-02 1  
## 5 1 Bd-Rv 0.9658986 2.364628e-02 1  
## 6 2 Bd-Rv 1.0000000 2.968210e-06 1  
## 7 3 Bd-Rv 0.3953450 1.028446e-01 1  
## 8 4 Bd-Rv 0.9839377 1.725382e-02 1  
## 9 1 Rv-Bd 0.9519832 3.174231e-02 1  
## 10 2 Rv-Bd 1.0000000 4.240509e-06 1  
## 11 3 Rv-Bd 0.3139700 9.581011e-02 1  
## 12 4 Rv-Bd 0.9772097 2.396191e-02 1

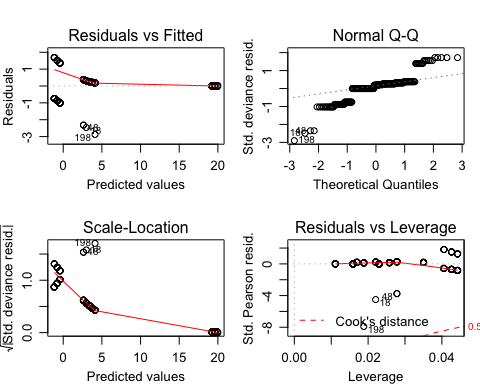


## Part 3d: Bd: model checks

Here I check the two best models, looking at the estimates of the coefficients using summary(model) and the diagnostic plots using plot(model)

### Maximal Model: Species + Treamtent (no interaction)

##   
## Call:  
## glm(formula = Bd.endpoint.status ~ Treatment + ExperimentNo,   
## family = binomial, data = Bd.status)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.87447 0.00007 0.17996 0.30114 1.68776   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 2.6197 0.6629 3.952 7.76e-05 \*\*\*  
## TreatmentBd-Rv 0.7240 0.6103 1.186 0.235   
## TreatmentRv-Bd 0.3673 0.6087 0.603 0.546   
## ExperimentNo2 16.6135 1379.7561 0.012 0.990   
## ExperimentNo3 -3.7686 0.6628 -5.686 1.30e-08 \*\*\*  
## ExperimentNo4 0.7714 1.1742 0.657 0.511   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 223.83 on 221 degrees of freedom  
## Residual deviance: 106.76 on 216 degrees of freedom  
## AIC: 118.76  
##   
## Number of Fisher Scoring iterations: 18

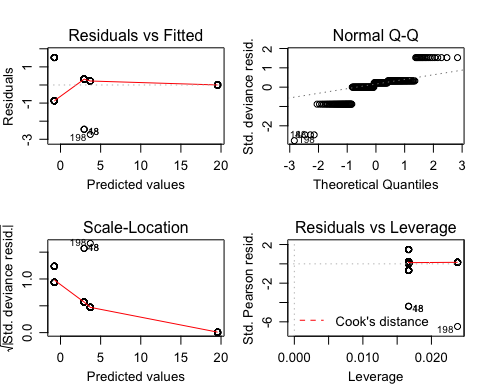


The diagnostic plots for glm(Bd.endpoint.status ~ Treatment + ExperimentNo) aren’t ideal

* residual vs. fitted shows patterning meaning the variance is non-consistent
* residual vs. leverage also shows patterning suggesting certain data points have strong influence

### Simplified Model: Species

##   
## Call:  
## glm(formula = Bd.endpoint.status ~ ExperimentNo, family = binomial,   
## data = Bd.status)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.73411 0.00008 0.21953 0.32029 1.51651   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 2.9444 0.5923 4.971 6.67e-07 \*\*\*  
## ExperimentNo2 16.6216 1388.3372 0.012 0.990   
## ExperimentNo3 -3.7136 0.6541 -5.677 1.37e-08 \*\*\*  
## ExperimentNo4 0.7691 1.1727 0.656 0.512   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 223.83 on 221 degrees of freedom  
## Residual deviance: 108.19 on 218 degrees of freedom  
## AIC: 116.19  
##   
## Number of Fisher Scoring iterations: 18



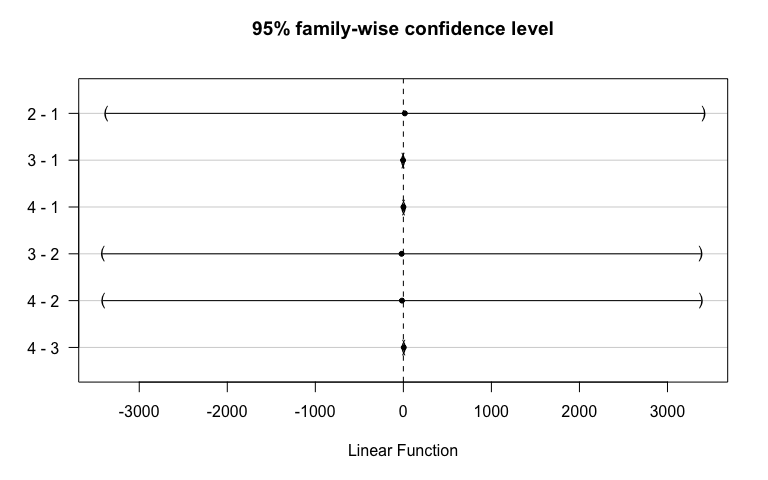
The diagnostic plots for glm(Bd.endpoint.status ~ ExperimentNo) aren’t much better

* residual vs. fitted shows patterning meaning the variance is non-consistent
* residual vs. leverage again certain data points have strong influence

# Part 4: Bd: post-hoc tests

Tukey’s honest significant difference How to interpret the plots “Comparisons having intervals that do not overlap the vertical dashed line are significanty different. The vertical dashed line indicates no difference between the mean values for the factor-level comparisons indicated on the y-axis.” ~ R BOOK

##   
## Simultaneous Tests for General Linear Hypotheses  
##   
## Multiple Comparisons of Means: Tukey Contrasts  
##   
##   
## Fit: glm(formula = Bd.endpoint.status ~ ExperimentNo, family = binomial,   
## data = Bd.status)  
##   
## Linear Hypotheses:  
## Estimate Std. Error z value Pr(>|z|)   
## 2 - 1 == 0 16.6216 1388.3372 0.012 1.000   
## 3 - 1 == 0 -3.7136 0.6541 -5.677 < 1e-05 \*\*\*  
## 4 - 1 == 0 0.7691 1.1727 0.656 0.893   
## 3 - 2 == 0 -20.3352 1388.3371 -0.015 1.000   
## 4 - 2 == 0 -15.8525 1388.3375 -0.011 1.000   
## 4 - 3 == 0 4.4827 1.0495 4.271 6.87e-05 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
## (Adjusted p values reported -- single-step method)



**Conclusion** the summary suggests that *Rana* are significantly different from *Bufo* and *Alytes* but the plot doesn’t show this. In fact, the plot suggests nothing is significant.

# 

# Part 5: Rv: : apply models to endpoint infection status

Here I create a dataframe with the Bd-only treatment group removed (as they have never been exposed to Rv) and without *Alytes muletensis* babies as they don’t have Rv tratment groups.

Rv.status <- data.endpoint %>%  
 filter(!Treatment=="Bd") %>%  
 filter(!ExperimentNo=='5') %>%   
 mutate(ExperimentNo = as.factor(ExperimentNo)) %>%   
 dplyr::select(ID, Species, ExperimentNo, Scenario, Treatment, Rv.MCPendpoint.status, Rv.endpoint.load)   
droplevels(Rv.status)

## Part 5a: Rv: model selection

The four models:

Rv.status1 <- glm(Rv.MCPendpoint.status ~ Treatment \* ExperimentNo, data=Rv.status, family=binomial)  
  
Rv.status2 <- glm(Rv.MCPendpoint.status ~ Treatment + ExperimentNo, data=Rv.status, family=binomial)  
  
Rv.status3 <- glm(Rv.MCPendpoint.status ~ Treatment, data=Rv.status, family=binomial)  
  
Rv.status4 <- glm(Rv.MCPendpoint.status ~ ExperimentNo, data=Rv.status, family=binomial)  
  
Rv.status.N <- glm(Rv.MCPendpoint.status ~ 1, data=Rv.status, family=binomial)

## Part 5b: Rv: model comparison

anova(Rv.status1, Rv.status2, test="Chisq") # start by comparing the interaction terms

## Analysis of Deviance Table  
##   
## Model 1: Rv.MCPendpoint.status ~ Treatment \* ExperimentNo  
## Model 2: Rv.MCPendpoint.status ~ Treatment + ExperimentNo  
## Resid. Df Resid. Dev Df Deviance Pr(>Chi)  
## 1 210 133.60   
## 2 216 139.25 -6 -5.6442 0.4642

anova(Rv.status2, Rv.status3, test="Chisq") # compares Trt and Species to just Trt

## Analysis of Deviance Table  
##   
## Model 1: Rv.MCPendpoint.status ~ Treatment + ExperimentNo  
## Model 2: Rv.MCPendpoint.status ~ Treatment  
## Resid. Df Resid. Dev Df Deviance Pr(>Chi)   
## 1 216 139.25   
## 2 219 146.02 -3 -6.7695 0.07962 .  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

anova(Rv.status2, Rv.status4, test="Chisq") # compares Trt and Species to just Species

## Analysis of Deviance Table  
##   
## Model 1: Rv.MCPendpoint.status ~ Treatment + ExperimentNo  
## Model 2: Rv.MCPendpoint.status ~ ExperimentNo  
## Resid. Df Resid. Dev Df Deviance Pr(>Chi)   
## 1 216 139.25   
## 2 218 158.13 -2 -18.887 7.921e-05 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

anova(Rv.status3, Rv.status.N, test="Chisq") # compares Trt model to null model

## Analysis of Deviance Table  
##   
## Model 1: Rv.MCPendpoint.status ~ Treatment  
## Model 2: Rv.MCPendpoint.status ~ 1  
## Resid. Df Resid. Dev Df Deviance Pr(>Chi)   
## 1 219 146.02   
## 2 221 164.34 -2 -18.328 0.0001048 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

anova(Rv.status4, Rv.status.N, test="Chisq") # compares Species model to null model

## Analysis of Deviance Table  
##   
## Model 1: Rv.MCPendpoint.status ~ ExperimentNo  
## Model 2: Rv.MCPendpoint.status ~ 1  
## Resid. Df Resid. Dev Df Deviance Pr(>Chi)  
## 1 218 158.13   
## 2 221 164.34 -3 -6.2102 0.1018

ANOVA 1: suggests we should reject the more complex model with interaction terms (Treatment \* ExperimentNo) in favour for the model with just the terms (pvalue = 0.4642)

ANOVA 2: suggests we should reject the more complex model (Treatment + ExperimentNo) and choose the model with just Treatment (pvalue = 0.07962)

ANOVA 3: suggests we should favour the more complex model (Treatment + ExperimentNo) over the model with ExperimentNo. (pvalue = < .001)

ANOVA 4: suggests we should favour the more complex model (Treatment) over the null model (pvalue = < .001) so Treatment leads to significantly improved fit

ANOVA 5: suggests we should reject the more complex model (ExperimentNo) and choose the null model

**Conclusion**: we should choose the model with just Treatment

## 

## Part 5c: Rv: model fit

## ID Species ExperimentNo Scenario Treatment Rv.MCPendpoint.status  
## 1 A1.1 Bb 1 Coinfection Rv-Bd 0  
## 2 A1.2 Bb 1 Coinfection Rv-Bd 1  
## 3 A1.3 Bb 1 Coinfection Rv-Bd 0  
## 4 A1.4 Bb 1 Coinfection Rv-Bd 1  
## 5 A1.5 Bb 1 Coinfection Rv-Bd 0  
## 6 A1.6 Bb 1 Coinfection Rv-Bd 1  
## Rv.endpoint.load bi.glm  
## 1 NA 0.2567568  
## 2 1376.3337 0.2567568  
## 3 NA 0.2567568  
## 4 913.5792 0.2567568  
## 5 NA 0.2567568  
## 6 811.7093 0.2567568

?? **QUESTION** It looks like the model is struggling to predict the probability of Rv infection status accurately ((**edit** *… in the Rv only treatment group. I agree that the predictions for both the coinfection groups match the raw data*))

## Part 5d: Rv: model plotting

## ExperimentNo Treatment fit se.fit residual.scale  
## 1 1 Bd-Rv 0.06756757 0.02917835 1  
## 2 2 Bd-Rv 0.06756757 0.02917835 1  
## 3 3 Bd-Rv 0.06756757 0.02917835 1  
## 4 4 Bd-Rv 0.06756757 0.02917835 1  
## 5 1 Rv 0.04054055 0.02292120 1  
## 6 2 Rv 0.04054055 0.02292120 1  
## 7 3 Rv 0.04054055 0.02292120 1  
## 8 4 Rv 0.04054055 0.02292120 1  
## 9 1 Rv-Bd 0.25675676 0.05078209 1  
## 10 2 Rv-Bd 0.25675676 0.05078209 1  
## 11 3 Rv-Bd 0.25675676 0.05078209 1  
## 12 4 Rv-Bd 0.25675676 0.05078209 1

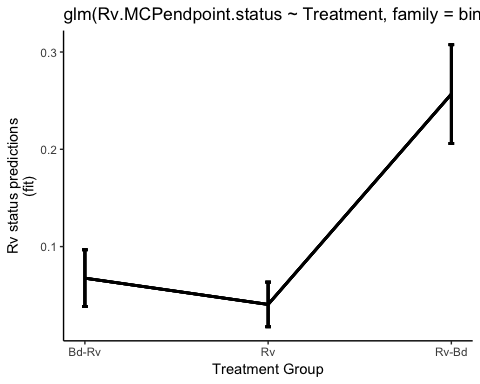


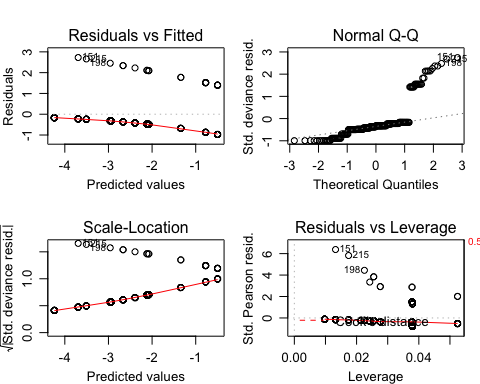
Fig. 3. Probability of Rv infection status (GE) predicted by model

## Part 5e: Rv: model checks

Here I check the two best models, looking at the estimates of the coefficients using summary(model) and the diagnostic plots using plot(model)

### Maximal Model: Species + Treamtent (no interaction)

##   
## Call:  
## glm(formula = Rv.MCPendpoint.status ~ Treatment + ExperimentNo,   
## family = binomial, data = Rv.status)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -0.9738 -0.4766 -0.3213 -0.2217 2.7269   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -2.3911 0.5616 -4.258 2.07e-05 \*\*\*  
## TreatmentRv -0.5467 0.7545 -0.725 0.46869   
## TreatmentRv-Bd 1.6181 0.5431 2.979 0.00289 \*\*   
## ExperimentNo2 0.2731 0.5241 0.521 0.60234   
## ExperimentNo3 -1.3023 0.7170 -1.816 0.06931 .   
## ExperimentNo4 -0.5684 0.6675 -0.851 0.39453   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 164.34 on 221 degrees of freedom  
## Residual deviance: 139.25 on 216 degrees of freedom  
## AIC: 151.25  
##   
## Number of Fisher Scoring iterations: 6

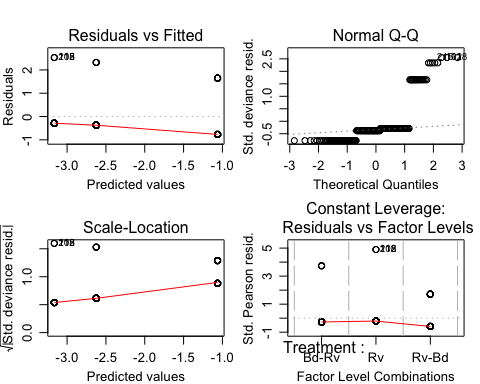


The diagnostic plots for glm(Rv.MCPendpoint.status ~ Treatment + ExperimentNo) again aren’t ideal

* QQ plot is iffy
* residual vs. fitted shows patterning meaning the variance is non-consistent
* residual vs. leverage also shows patterning suggesting certain data points have strong influence

### Simplified Model: Treatment

##   
## Call:  
## glm(formula = Rv.MCPendpoint.status ~ Treatment, family = binomial,   
## data = Rv.status)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -0.7704 -0.3741 -0.3741 -0.2877 2.5320   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -2.6247 0.4631 -5.667 1.45e-08 \*\*\*  
## TreatmentRv -0.5394 0.7495 -0.720 0.47172   
## TreatmentRv-Bd 1.5618 0.5341 2.924 0.00346 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 164.34 on 221 degrees of freedom  
## Residual deviance: 146.02 on 219 degrees of freedom  
## AIC: 152.02  
##   
## Number of Fisher Scoring iterations: 5

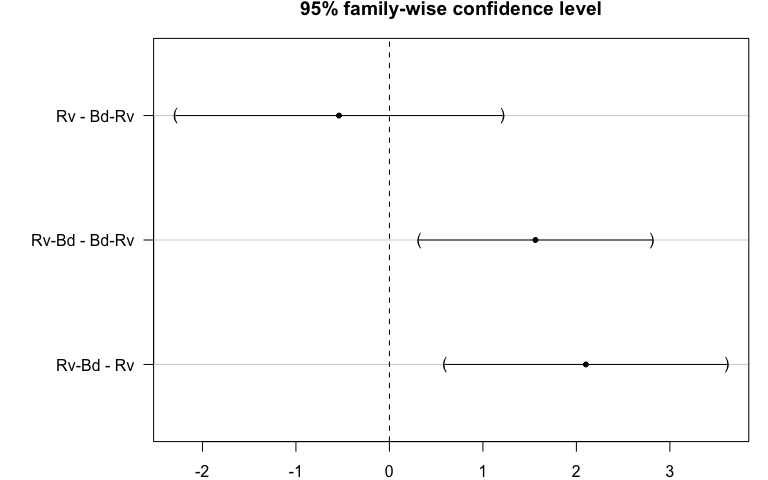


The diagnostic plots for glm(Rv.MCPendpoint.status ~ Treatment) aren’t much better

# Part 6: Rv: post-hoc tests

Tukey’s honest significant difference

##   
## Simultaneous Tests for General Linear Hypotheses  
##   
## Multiple Comparisons of Means: Tukey Contrasts  
##   
##   
## Fit: glm(formula = Rv.MCPendpoint.status ~ Treatment, family = binomial,   
## data = Rv.status)  
##   
## Linear Hypotheses:  
## Estimate Std. Error z value Pr(>|z|)   
## Rv - Bd-Rv == 0 -0.5394 0.7495 -0.720 0.74764   
## Rv-Bd - Bd-Rv == 0 1.5618 0.5341 2.924 0.00928 \*\*  
## Rv-Bd - Rv == 0 2.1012 0.6466 3.250 0.00310 \*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
## (Adjusted p values reported -- single-step method)



**Conclusion**

* Rv vs. Bd-Rv is not significant
* Rv-Bd vs. Bd-Rv there is significant difference between the coinfection groups
* Rv-Bd vs. Rv is significant

?? **QUESTION** is this highlighting the timing/sequence of dose again ??