R Course: lesson4

Cédric HASSEN-KHODJA

16 juin 2016

Introduction

I analyzed alien species data from two "Star Trek" series, "Star Trek: The Original Series" and "Star Trek: The Next Generation". Specifically, I looked at whether series ("The Original Series", "The Next Generation") and species alignment to the Enterprise (foe, friend) could predict whether the species was classified as likely to become extinct in the near future or not. Note, in the classification for this analysis only species with a classification of "least concerned" in a more nuanced classification system, were labeled as "not likely", the rest were labeled as "likely".

Results

I tested for if an alien species' likelihood of becoming extinct could be predicted by the series in which the species appeared and whether the species was a friend or a foe. Initial visual examination of the data suggests that there is an interaction, where likelihood of becoming extinct for friends or foes is flipped for each series.
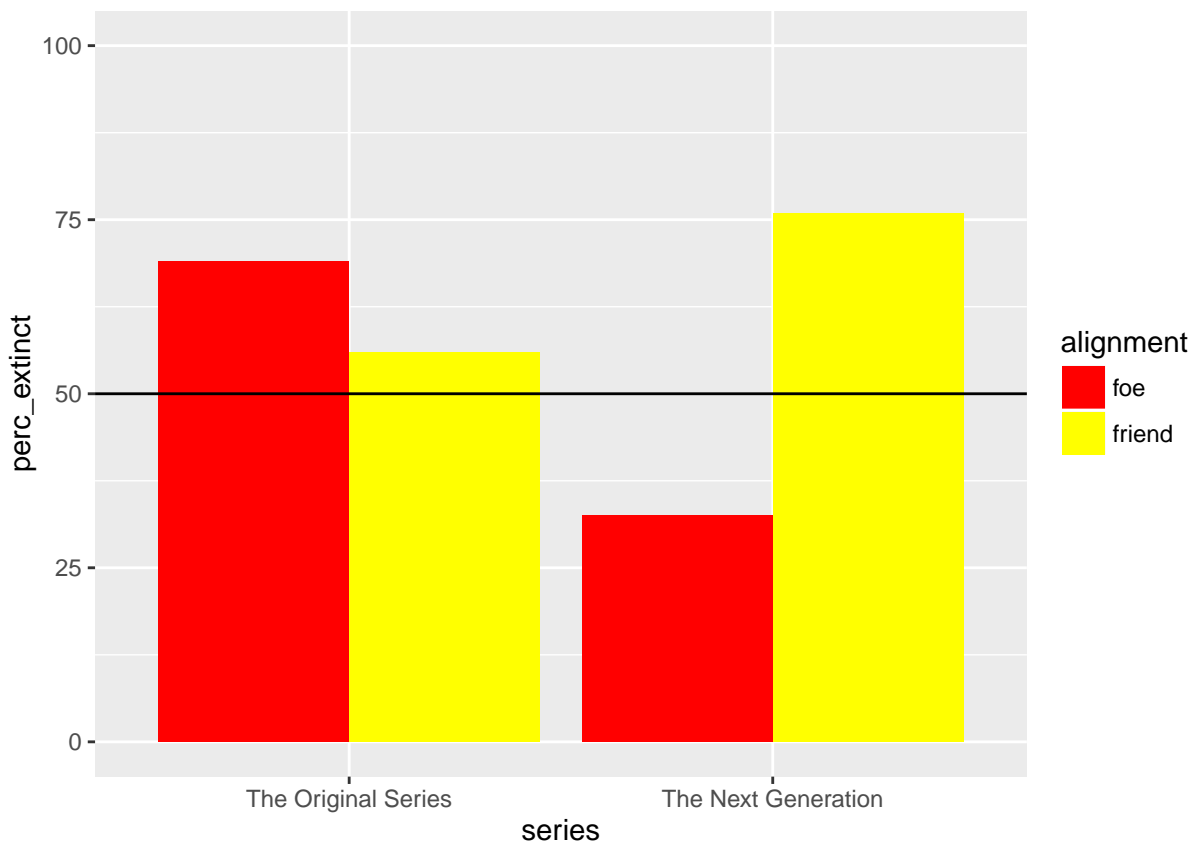
To test this effect I ran a logistic regression with "not likely to become extinct" (0) or "likely to become extinct" (1) as the dependent variable and series and alignment as independent variables. There was a significant effect of series and a significant interaction of series and alignment.

```
##
## Call:
## glm(formula = extinct ~ series * alignment, family = "binomial",
##     data = data_stats)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.6882  -0.8876   0.7418   0.8620   1.4981
##
## Coefficients:
##                          Estimate Std. Error z value Pr(>|z|)
## (Intercept)                0.7985     0.4014   1.989  0.04666 *
## seriestng                 -1.5267     0.5167  -2.955  0.00313 **
## alignmentfriend           -0.5621     0.5295  -1.062  0.28845
## seriestng:alignmentfriend  2.4403     0.6750   3.615  0.00030 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 247.30  on 184  degrees of freedom
## Residual deviance: 224.02  on 181  degrees of freedom
## AIC: 232.02
##
## Number of Fisher Scoring iterations: 4
```

To better understand the interaction in my model, I releveled my variables to see if effects changed. First I releveled the variable series to "tng". Now there is a significant effect of alignment, with friends being more likely to become extinct than foes, but specifically for "The Next Generation". There was no effect for "The Original Series" in the previous model.

```
##
## Call:
## glm(formula = extinct ~ relevel(series, "tng") * alignment, family = "binomial",
##     data = data_stats)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.6882  -0.8876   0.7418   0.8620   1.4981
##
## Coefficients:
##                                      Estimate Std. Error z value
## (Intercept)                           -0.7282     0.3254  -2.238
## relevel(series, "tng")tos              1.5267     0.5167   2.955
## alignmentfriend                        1.8781     0.4186   4.487
## relevel(series, "tng")tos:alignmentfriend -2.4403     0.6750  -3.615
##                                      Pr(>|z|)
## (Intercept)                           0.02524 *
```

```
## relevel(series, "tng")tos                   0.00313 **
## alignmentfriend                             7.23e-06 ***
## relevel(series, "tng")tos:alignmentfriend  0.00030 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 247.30  on 184  degrees of freedom
## Residual deviance: 224.02  on 181  degrees of freedom
## AIC: 232.02
##
## Number of Fisher Scoring iterations: 4
```

I also releveled the variable alignment to "friend" to see if my effect of series persists or goes away. As one can see I still have an effect of series, but now in the opposite direction. So while foes were significantly more likely to *not* become extinct in "The Next Generation" than "The Original Series", the reverse is true for friends.

extinct_seriesxalignment_friend.glm_sum

```
##
## Call:
## glm(formula = extinct ~ series * relevel(alignment, "friend"),
##     family = "binomial", data = data_stats)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.6882  -0.8876   0.7418   0.8620   1.4981
##
## Coefficients:
##                                          Estimate Std. Error z value
## (Intercept)                                0.2364     0.3454   0.684
## seriestng                                  0.9135     0.4343   2.104
## relevel(alignment, "friend")foe            0.5621     0.5295   1.062
## seriestng:relevel(alignment, "friend")foe -2.4403     0.6750  -3.615
##                                          Pr(>|z|)
## (Intercept)                                0.4937
## seriestng                                  0.0354 *
## relevel(alignment, "friend")foe            0.2884
## seriestng:relevel(alignment, "friend")foe  0.0003 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 247.30  on 184  degrees of freedom
## Residual deviance: 224.02  on 181  degrees of freedom
## AIC: 232.02
##
## Number of Fisher Scoring iterations: 4
```

Conclusion

The likelihood of a "Star Trek" species to be headed towards extinction was affected both by the series where the species was portrayed and whether the species was a friend or a foe to the Enterprise. There was a significant interaction, with alignment not mattering in "The Original Series" but having an effect in "The Next Generation". This may have to do with the social issues and politics at the time each series was filmed, "The Original Series" 1966-1969 versus "The Next Generation" 1987-1994.