

P8131 HW5

Brian Jo Hsuan Lee

3/15/2022

Load packages

```
library(tidyverse)
```

Problem 1: Crab Satellite Count

Load and tidy data

```
# txt file read in using read_delim() and cleaned accordingly
crab_data =
  read_delim("HW5-crab.txt", delim = "\t") %>%
  mutate(
    number = str_trim(number, side = c("both"))
  ) %>%
  separate(number, c("number", "C", "S", "W", "Wt", "Sa"), sep = " +") %>%
  mutate(
    across(where(is.character), as.numeric)
  )
```

```
## Rows: 173 Columns: 6
```

```
## -- Column specification -----
```

```
## Delimiter: "\t"
```

```
## chr (1): number
```

```
## lgl (5): C, S, W, Wt, Sa
```

```
##
```

```
## i Use `spec()` to retrieve the full column specification for this data.
```

```
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

a) **Fit a simple Poisson model, check the goodness of fit and interpret the model**

```
# m1 model fit
crab_m1 = glm(Sa ~ W, family = poisson, data = crab_data)
summary(crab_m1)
```

```
##
```

```
## Call:
```

```
## glm(formula = Sa ~ W, family = poisson, data = crab_data)
```

```
##
```

```
## Deviance Residuals:
```

```
##      Min       1Q   Median       3Q      Max
```

```
## -2.8526  -1.9884  -0.4933   1.0970   4.9221
```

```
##
```

```
## Coefficients:
```

```
##           Estimate Std. Error z value Pr(>|z|)
## (Intercept) -3.30476    0.54224  -6.095  1.1e-09 ***
## W           0.16405    0.01997   8.216  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
## Null deviance: 632.79  on 172  degrees of freedom
## Residual deviance: 567.88  on 171  degrees of freedom
## AIC: 927.18
##
## Number of Fisher Scoring iterations: 6
# both deviance residual and peason's residual goodness of fit tests,
# with df = 173 observations - 2 predictors = 171
crab_m1_deviance = crab_m1$deviance
crab_m1_deviance_pval = 1 - pchisq(crab_m1_deviance, 171)
crab_m1_pchisq = sum(residuals(crab_m1, 'pearson')^2)
crab_m1_pchisq_pval = 1 - pchisq(crab_m1_pchisq, 171)

ifelse(crab_m1_deviance_pval > 0.05 | crab_m1_pchisq_pval > 0.05,
       'Failed to reject the null, since no significant evidence suggest the poisson fit is not good',
       'Reject the null with significant data suggesting the poisson fit is not good')

## [1] "Reject the null with significant data suggesting the poisson fit is not good"
```

b) Fit a Poisson model with 2 predictors, compare it with the previous model and interpret it

```
# m2 model fit
crab_m2 = glm(Sa ~ W + Wt, family = poisson, data = crab_data)
summary(crab_m2)

##
## Call:
## glm(formula = Sa ~ W + Wt, family = poisson, data = crab_data)
##
## 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    0.89929  -1.436  0.15091
## W           0.04590    0.04677   0.981  0.32640
## Wt          0.44744    0.15864   2.820  0.00479 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
## 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
# use chisq test and evaluate the nested models,
# with df = 171 m1 predictors - 170 m2 predictors = 1
m1_m2_stat = crab_m1$deviance - crab_m2$deviance
m1_m2_pval = 1 - pchisq(m1_m2_stat, df = 171-170)
ifelse(m1_m2_pval > 0.05,
      'Failed to reject the null, since no significant evidence suggest the larger model has a better :
      'Reject the null with significant evidence suggesting the larger model fits the data better')

## [1] "Reject the null with significant evidence suggesting the larger model fits the data better"

```