DS 2021 Jelly Analyses

Mara Bohm and Naomi Lubkin

15/12/2021

Loading In Packages and Data

```r
library(patchwork)
library(car)
library(performance)
library(DHARMa)
library(fitdistrplus)
library(gamlss)
library(tidyverse)

polyp_data <- read_csv("/Users/marabohm/Github/Jelly-Data/data/ds2021_polyp_data.csv")
```

Cleaning the Data

```r
polyp_clean <- polyp_data %>%
    rename(collection_day = "Data collection day", jar_code = "Jar Code (ex. E1)",
        treatment = Chemical, num_elongated = "Elongation?", num_ruffled = "Ruffled?",
        num_asexual_buds = "Asexual Repro?", total_num = Total) %>%
    mutate(jar_code = as.factor(jar_code), treatment = as.factor(treatment),
        treatment = fct_relevel(treatment, "Control", "Caffeine", "Estradiol",
            "Combo")) %>%
    dplyr::select(collection_day, jar_code, treatment, num_elongated, num_ruffled,
        num_asexual_buds, total_num)
```

Polyp Elongation

Step 1: Fitting Distributions for Elongation Data

```r
fitDist(num_elongated, data = polyp_clean, type = "counts", try.gamlss = T)
```

```
##   |                                                                    |
```

```
##
## Family:  c("NBII", "Negative Binomial type II")
## Fitting method: "nlminb"
##
## Call:  gamlssML(formula = y, family = DIST[i])
##
## Mu Coefficients:
## [1]   2.504
## Sigma Coefficients:
## [1]   1.662
##
##  Degrees of Freedom for the fit: 2 Residual Deg. of Freedom    58
## Global Deviance:       410.587
##              AIC:      414.587
##              SBC:      418.776
```

```
# best fit: Negative Binomial type II (AIC = 414.587)
```

```
# visualizing this fit
histDist(polyp_clean$num_elongated, "NBII", density = T, main = "Count Polyp Elongation
        Compared to Negative Binomial II Distibution",
    xlab = "Number Elongated")
```


Count Polyp Elongation
Compared to Negative Binomial II Distibution
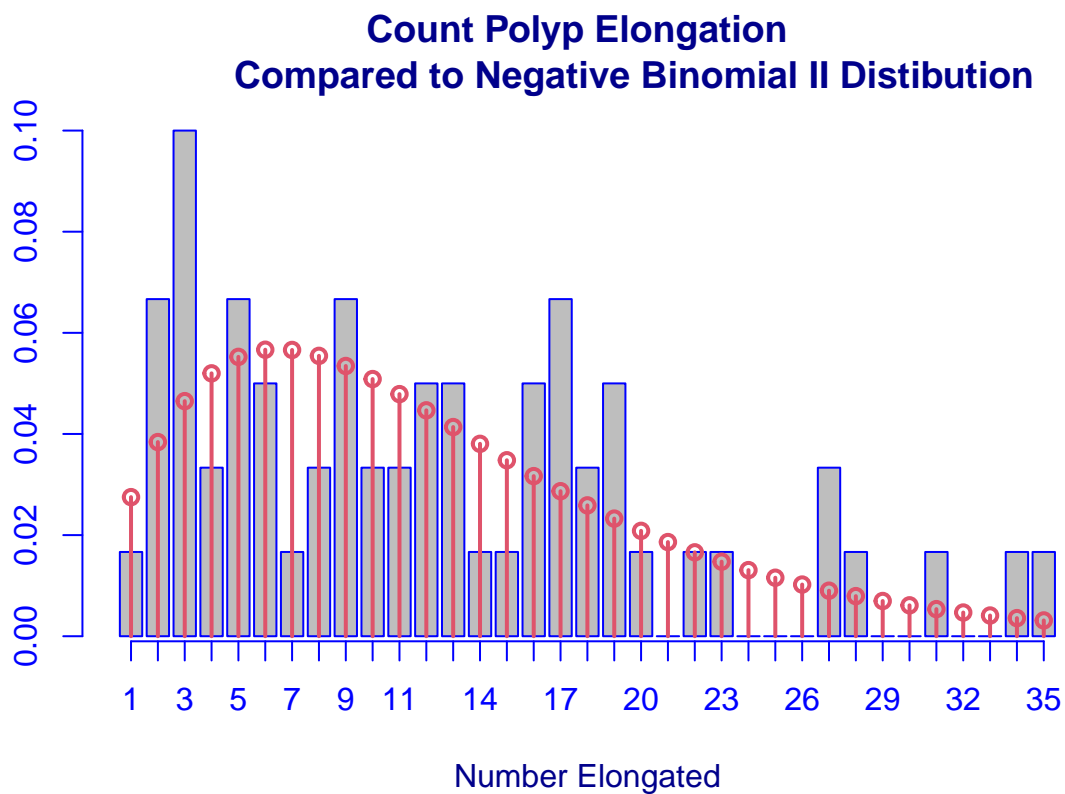
Step 2: Building the Model WITH Total As A Random Effect

```
mod_WITHtotal_elongation <- gamlss(num_elongated ~ treatment * collection_day +
    re(random = ~1 | jar_code) + re(random = ~1 | total_num), family = NBII(),
    method = RS(), data = polyp_clean, control = gamlss.control(n.cyc = 200))
```

Step 3: Building the Model WITHOUT Total As A Random Effect

```
mod_NOtotal_elongation <- gamlss(num_elongated ~ treatment * collection_day +
    re(random = ~1 | jar_code), family = NBII(), method = RS(), data = polyp_clean,
    control = gamlss.control(n.cyc = 200))
```

Step 4: Comparing The With and Without Total Models Using GAIC

```
GAIC(mod_NOtotal_elongation, mod_WITHtotal_elongation)
```

```
##                             df     AIC
## mod_WITHtotal_elongation 8.35727 333.33
## mod_NOtotal_elongation   9.35727 335.33
```

Step 5: Models Are Effectively The Same - Summarizing With Total Model

```
summary(mod_WITHtotal_elongation)
```

```
## ******************************************************************
## Family:  c("NBII", "Negative Binomial type II")
##
## Call:  gamlss(formula = num_elongated ~ treatment * collection_day +
##     re(random = ~1 | jar_code) + re(random = ~1 | total_num),
##     family = NBII(), data = polyp_clean, method = RS(),
##     control = gamlss.control(n.cyc = 200))
##
## Fitting method: RS()
##
## -------------------------------------------------------------------
## Mu link function:  log
## Mu Coefficients:
##                                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)                      1.09720    0.28047   3.912 0.000269 ***
## treatmentCaffeine               -0.17935    0.38941  -0.461 0.647033
## treatmentEstradiol               0.03055    0.37398   0.082 0.935219
## treatmentCombination             0.11280    0.36307   0.311 0.757292
## collection_day                   0.31578    0.07282   4.336 6.74e-05 ***
## treatmentCaffeine:collection_day 0.11929    0.09911   1.204 0.234253
## treatmentEstradiol:collection_day 0.10750   0.09556   1.125 0.265833
## treatmentCombination:collection_day 0.14295  0.09252   1.545 0.128455
```

3

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## -----------------------------------------------------------------
## Sigma link function:  log
## Sigma Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -1.4610     0.9872   -1.48    0.145
##
## -----------------------------------------------------------------
## NOTE: Additive smoothing terms exist in the formulas:
##  i) Std. Error for smoothers are for the linear effect only.
## ii) Std. Error for the linear terms maybe are not accurate.
## -----------------------------------------------------------------
## No. of observations in the fit:  60
## Degrees of Freedom for the fit:  8.35727
##       Residual Deg. of Freedom:  51.64273
##                      at cycle:  43
##
## Global Deviance:     316.6155
##             AIC:     333.33
##             SBC:     350.833
## *****************************************************************
```

Polyp Ruffling

Step 1: Fitting Distributions for Ruffling Data

```
fitDist(num_ruffled, data = polyp_clean, type = "counts", try.gamlss = T)
# best fit: Negative Binomial type II (AIC = 444.109)

# model was having difficulty converging with NBII, so tried some
# other fits:

descdist(polyp_clean$num_ruffled)


# Cullen and Frey plot indicates Lognormal, Gamma and Weibull
# distributions may be a good fit, so we compare to each other using
# Generalized AIC:

b_NB2 <- histDist(polyp_clean$num_ruffled, "NBII", density = T)


b_lNO <- histDist(polyp_clean$num_ruffled, "LOGNO", density = T)


b_GA <- histDist(polyp_clean$num_ruffled, "GA", density = T)


b_WEI <- histDist(polyp_clean$num_ruffled, "WEI", density = T)
```
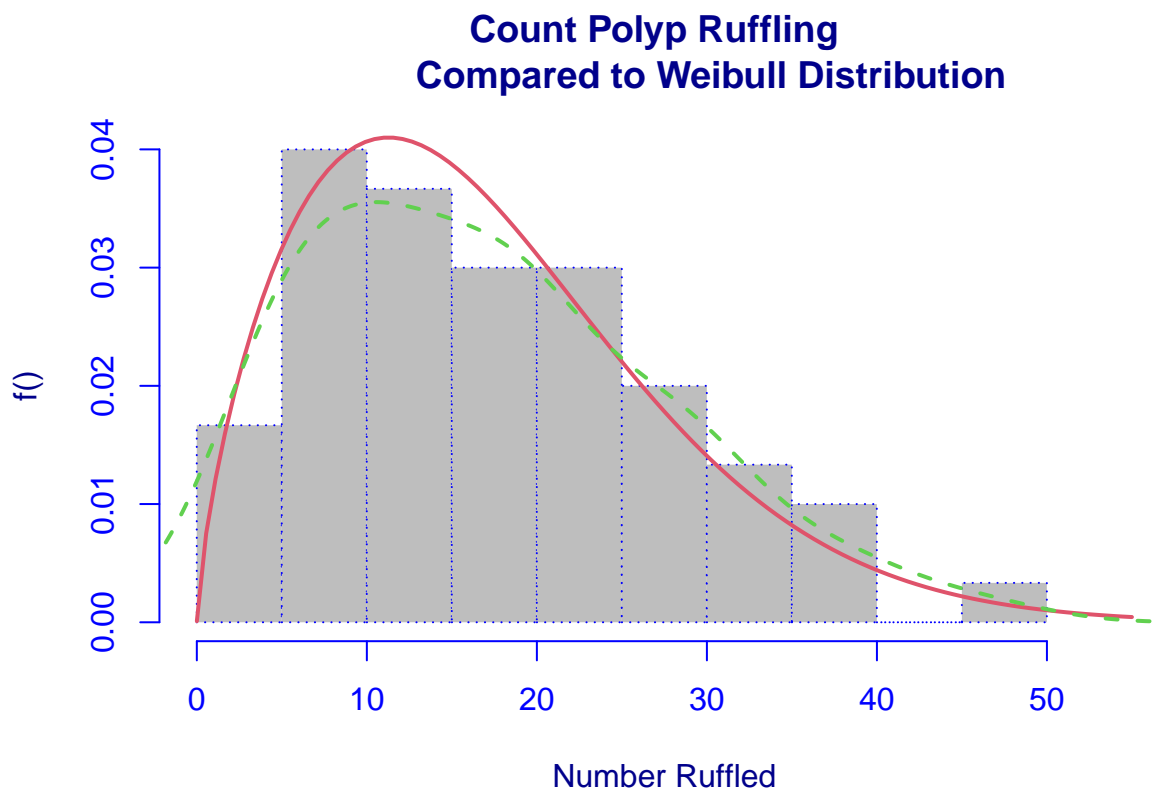
```
GAIC(b_lNO, b_GA, b_WEI, b_NB2)
```

```
##        df      AIC
## b_WEI   2 442.8371
## b_NB2   2 444.1095
## b_GA    2 444.1204
## b_lNO   2 451.8062
```

```
# best fit: Weibull Distribution (AIC = 442.837)
```

```
# visualizing this fit
histDist(polyp_clean$num_ruffled, "WEI", density = T, main = "Count Polyp Ruffling
         Compared to Weibull Distribution",
    xlab = "Number Ruffled")
```


Count Polyp Ruffling
Compared to Weibull Distribution

```
##
## Family:  c("WEI", "Weibull")
## Fitting method: "nlminb"
##
## Call:  gamlssML(formula = polyp_clean$num_ruffled, family = "WEI")
##
## Mu Coefficients:
## [1]  2.947
## Sigma Coefficients:
```

5

```
## [1]  0.529
##
##  Degrees of Freedom for the fit: 2 Residual Deg. of Freedom    58
## Global Deviance:      438.837
##             AIC:      442.837
##             SBC:      447.026
```

Step 2: Building the Model WITH Total As A Random Effect

```
mod_WITHtotal_ruffled <- gamlss(num_ruffled ~ treatment * collection_day +
    re(random = ~1 | jar_code) + re(random = ~1 | total_num), family = WEI(),
    data = polyp_clean, control = gamlss.control(n.cyc = 60))
```

Step 3: Building the Model WITHOUT Total As A Random Effect

```
mod_NOtotal_ruffled <- gamlss(num_elongated ~ treatment * collection_day +
    re(random = ~1 | jar_code), family = WEI(), data = polyp_clean, control = gamlss.control(n.cyc = 20
```

Step 4: Comparing The With and Without Total Models Using GAIC

```
GAIC(mod_NOtotal_ruffled, mod_WITHtotal_ruffled)
```

```
##                            df      AIC
## mod_NOtotal_ruffled   13.49063 346.1313
## mod_WITHtotal_ruffled 14.64746 348.9679
```

Step 5: Models Are Effectively The Same - Summarizing With Total Model

```
summary(mod_WITHtotal_ruffled)
```

```
## *********************************************************************
## Family:  c("WEI", "Weibull")
##
## Call:  gamlss(formula = num_ruffled ~ treatment * collection_day +
##     re(random = ~1 | jar_code) + re(random = ~1 | total_num),
##     family = WEI(), data = polyp_clean, control = gamlss.control(n.cyc = 60))
##
## Fitting method: RS()
##
## -------------------------------------------------------------------
## Mu link function:  log
## Mu Coefficients:
##                                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)                     1.231058   0.161918   7.603 1.25e-09 ***
## treatmentCaffeine               0.471997   0.207627   2.273   0.0278 *
```

```
## treatmentEstradiol                  0.103466   0.224071   0.462   0.6465
## treatmentCombination               0.503957   0.224694   2.243   0.0298 *
## collection_day                     0.442230   0.050638   8.733 2.84e-11 ***
## treatmentCaffeine:collection_day  -0.063279   0.064249  -0.985   0.3299
## treatmentEstradiol:collection_day -0.008946   0.070142  -0.128   0.8991
## treatmentCombination:collection_day -0.055425 0.070382  -0.787   0.4351
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## ----------------------------------------------------------------------
## Sigma link function:  log
## Sigma Coefficients:
##            Estimate Std. Error t value Pr(>|t|)
## (Intercept)   1.5480     0.1053    14.7   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## ----------------------------------------------------------------------
## NOTE: Additive smoothing terms exist in the formulas:
##  i) Std. Error for smoothers are for the linear effect only.
## ii) Std. Error for the linear terms maybe are not accurate.
## ----------------------------------------------------------------------
## No. of observations in the fit:  60
## Degrees of Freedom for the fit:  14.64746
##       Residual Deg. of Freedom:  45.35254
##                      at cycle:  3
##
## Global Deviance:     319.673
##            AIC:     348.9679
##            SBC:     379.6448
## **********************************************************************
```

Polyp Asexual Budding

Step 1: Fitting Distributions for Buds Data

```
fitDist(num_asexual_buds, data = polyp_clean, type = "counts", try.gamlss = T)
```

```
##   |                                                                      |
##   Lapack routine dgesv: system is exactly singular: U[1,1] = 0
##   |                                                                      |=============
##   Lapack routine dgesv: system is exactly singular: U[2,2] = 0
##   |                                                                      |===================
```

```
##
## Family:  c("NBII", "Negative Binomial type II")
## Fitting method: "nlminb"
##
## Call:  gamlssML(formula = y, family = DIST[i])
##
## Mu Coefficients:
```

```
## [1]   2.041
## Sigma Coefficients:
## [1]   1.575
##
##  Degrees of Freedom for the fit: 2 Residual Deg. of Freedom    58
## Global Deviance:      368.321
##             AIC:      372.321
##             SBC:      376.509
```

```
# best fit: Negative Binomial type II (AIC = 372.321)
```

```
# visualizing this fit
histDist(polyp_clean$num_asexual_buds, "NBII", density = T, main = "Count Polyp Budding
          Compared to Negative Binomial II Distibution",
    xlab = "Number of Buds")
```
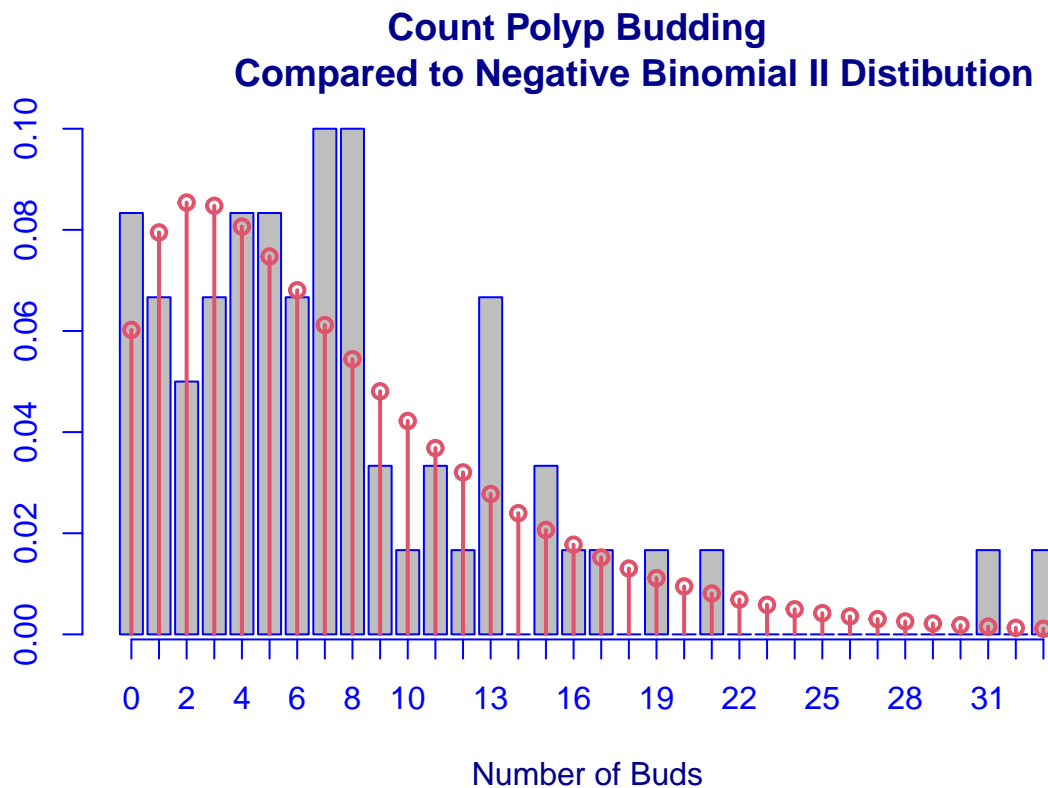

Step 2: Building the Model WITH Total As A Random Effect

```
mod_WITHtotal_buds <- gamlss(num_asexual_buds ~ treatment * collection_day +
    re(random = ~1 | jar_code) + re(random = ~1 | total_num), family = NBII(),
    data = polyp_clean, control = gamlss.control(n.cyc = 250))
```

8

Step 3: Building the Model WITHOUT Total As A Random Effect

```
mod_NOtotal_buds <- gamlss(num_elongated ~ treatment * collection_day +
    re(random = ~1 | jar_code), family = NBII(), data = polyp_clean, control = gamlss.control(n.cyc = 2
```

Step 4: Comparing The With and Without Total Models Using GAIC

```
GAIC(mod_NOtotal_buds, mod_WITHtotal_buds)
```

```
##                          df      AIC
## mod_WITHtotal_buds 11.04057 293.8201
## mod_NOtotal_buds    9.35727 335.3300
```

Step 5: Summarizing the Better WITH Total Model

```
summary(mod_WITHtotal_buds)
```

```
## *******************************************************************
## Family:  c("NBII", "Negative Binomial type II")
##
## Call:  gamlss(formula = num_asexual_buds ~ treatment * collection_day +
##      re(random = ~1 | jar_code) + re(random = ~1 | total_num),
##      family = NBII(), data = polyp_clean, control = gamlss.control(n.cyc = 250))
##
##
## Fitting method: RS()
##
## ---------------------------------------------------------------------
## Mu link function:  log
## Mu Coefficients:
##                                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)                      -0.54561    0.44010  -1.240  0.22097
## treatmentCaffeine                 0.79621    0.54887   1.451  0.15326
## treatmentEstradiol                0.17758    0.57850   0.307  0.76017
## treatmentCombination              1.66274    0.50878   3.268  0.00198 **
## collection_day                    0.60862    0.10430   5.835  4.2e-07 ***
## treatmentCaffeine:collection_day -0.08091    0.13120  -0.617  0.54027
## treatmentEstradiol:collection_day 0.06207    0.13640   0.455  0.65108
## treatmentCombination:collection_day -0.26110  0.12315  -2.120  0.03908 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## ---------------------------------------------------------------------
## Sigma link function:  log
## Sigma Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -2.526      2.645  -0.955    0.344
##
```

```
## ---------------------------------------------------------------------
## NOTE: Additive smoothing terms exist in the formulas:
##  i) Std. Error for smoothers are for the linear effect only.
## ii) Std. Error for the linear terms maybe are not accurate.
## ---------------------------------------------------------------------
## No. of observations in the fit:  60
## Degrees of Freedom for the fit:  11.04057
##        Residual Deg. of Freedom:  48.95943
##                         at cycle:  8
##
## Global Deviance:      271.739
##             AIC:      293.8201
##             SBC:      316.9429
## ******************************************************************
```

Polyp Total Number

Step 1: Fitting Distributions for Total Data

```r
fitDist(total_num, data = polyp_clean, type = "counts", try.gamlss = T)
```

```
##    |                                                                |
##    system is computationally singular: reciprocal condition number = 9.17006e-23
##    |                                                                |===================
```

```
##
## Family:  c("DEL", "Delaporte")
## Fitting method: "nlminb"
##
## Call:  gamlssML(formula = y, family = DIST[i])
##
## Mu Coefficients:
## [1]  3.208
## Sigma Coefficients:
## [1]  0.2898
## Nu Coefficients:
## [1]  0.5139
##
##  Degrees of Freedom for the fit: 3 Residual Deg. of Freedom    57
## Global Deviance:      430.754
##             AIC:      436.754
##             SBC:      443.038
```

```r
# best fit: delaporte (AIC = 436.754)
```

```r
# visualizing this fit
histDist(polyp_clean$total_num, "DEL", density = T, main = "Count Polyp Total
        Compared to Delaporte Distribution",
    xlab = "Number Polyps Total")
```

**Count Polyp Total
Compared to Delaporte Distribution**

Number Polyps Total
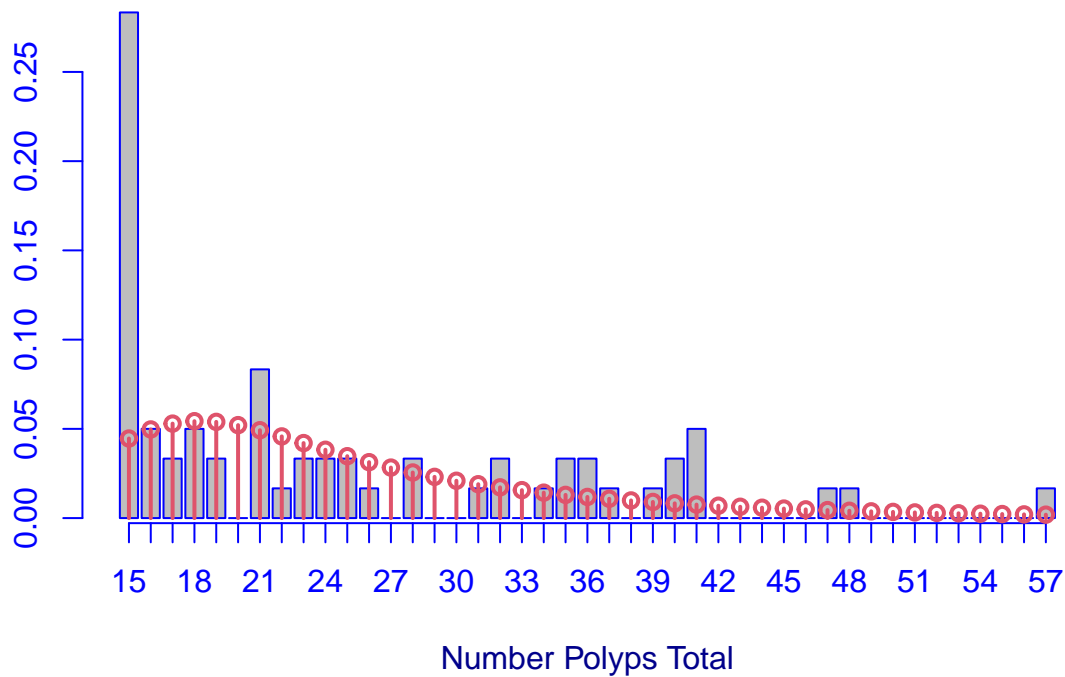
Step 2: Building the Model

```
mod_total <- gamlss(total_num ~ treatment * collection_day + re(random = ~1 |
    jar_code), family = DEL(), data = polyp_clean, control = gamlss.control(n.cyc = 60))
```

Step 3: Summarizing the Model

```
summary(mod_total)
```

```
## ******************************************************************
## Family:  c("DEL", "Delaporte")
##
## Call:
## gamlss(formula = total_num ~ treatment * collection_day + re(random = ~1 |
##     jar_code), family = DEL(), data = polyp_clean, control = gamlss.control(n.cyc = 60))
##
##
## Fitting method: RS()
##
## -----------------------------------------------------------------
## Mu link function:  log
```

```
## Mu Coefficients:
##                                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)                           2.364608   0.356013   6.642 2.33e-08 ***
## treatmentCaffeine                    -0.011063   0.522953  -0.021   0.9832
## treatmentEstradiol                    0.005354   0.469892   0.011   0.9910
## treatmentCombination                  0.048451   0.429514   0.113   0.9106
## collection_day                        0.223758   0.084778   2.639   0.0111 *
## treatmentCaffeine:collection_day      0.050239   0.128599   0.391   0.6977
## treatmentEstradiol:collection_day    -0.005437   0.111598  -0.049   0.9613
## treatmentCombination:collection_day   0.069381   0.105428   0.658   0.5135
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## ------------------------------------------------------------------
## Sigma link function:  log
## Sigma Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -36.04   12909.94  -0.003    0.998
##
## ------------------------------------------------------------------
## Nu link function:  logit
## Nu Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)        0      12910       0        1
##
## ------------------------------------------------------------------
## NOTE: Additive smoothing terms exist in the formulas:
##  i) Std. Error for smoothers are for the linear effect only.
## ii) Std. Error for the linear terms may not be reliable.
## ------------------------------------------------------------------
## No. of observations in the fit:  60
## Degrees of Freedom for the fit:  12.73762
##       Residual Deg. of Freedom:  47.26238
##                      at cycle:  5
##
## Global Deviance:     318.8629
##             AIC:     344.3381
##             SBC:     371.0151
## ******************************************************************
```

Is There A Difference Between Caffeine and Combo in Ruffling?

Step 1: Reordering Ruffling Data To Compare to Caffeine

```r
comparisons_data <- polyp_data %>%
   rename(collection_day = "Data collection day", jar_code = "Jar Code (ex. E1)",
      treatment = Chemical, num_elongated = "Elongation?", num_ruffled = "Ruffled?",
      num_asexual_buds = "Asexual Repro?", total_num = Total) %>%
   mutate(jar_code = as.factor(jar_code), treatment = as.factor(treatment)) %>%
   dplyr::select(collection_day, jar_code, treatment, num_elongated, num_ruffled,
      num_asexual_buds, total_num)
```

Step 2: Building the Reordered Ruffled Model

```
mod_WITHtotal_ruffled_reordered <- gamlss(num_ruffled ~ treatment * collection_day +
    re(random = ~1 | jar_code) + re(random = ~1 | total_num), family = WEI(),
    data = comparisons_data, control = gamlss.control(n.cyc = 250))
```

Step 3: Summarizing the Model

```
summary(mod_WITHtotal_ruffled_reordered)
```

```
## *******************************************************************
## Family:  c("WEI", "Weibull")
##
## Call:  gamlss(formula = num_ruffled ~ treatment * collection_day +
##     re(random = ~1 | jar_code) + re(random = ~1 | total_num),
##     family = WEI(), data = comparisons_data, control = gamlss.control(n.cyc = 250))
##
##
## Fitting method: RS()
##
## ---------------------------------------------------------------------
## Mu link function:  log
## Mu Coefficients:
##                                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)                         1.703055   0.133629  12.745  < 2e-16 ***
## treatmentCombination                0.031960   0.204865   0.156   0.8767
## treatmentControl                   -0.471997   0.207627  -2.273   0.0278 *
## treatmentEstradiol                 -0.368531   0.204012  -1.806   0.0775 .
## collection_day                      0.378951   0.039966   9.482 2.48e-12 ***
## treatmentCombination:collection_day 0.007854   0.063151   0.124   0.9016
## treatmentControl:collection_day     0.063279   0.064249   0.985   0.3299
## treatmentEstradiol:collection_day   0.054333   0.062840   0.865   0.3918
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## ---------------------------------------------------------------------
## Sigma link function:  log
## Sigma Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   1.5480     0.1053    14.7   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## ---------------------------------------------------------------------
## NOTE: Additive smoothing terms exist in the formulas:
##  i) Std. Error for smoothers are for the linear effect only.
## ii) Std. Error for the linear terms maybe are not accurate.
## ---------------------------------------------------------------------
## No. of observations in the fit:  60
## Degrees of Freedom for the fit:  14.64746
```

13

```
##       Residual Deg. of Freedom:  45.35254
##                      at cycle:  3
##
## Global Deviance:      319.673
##            AIC:       348.9679
##            SBC:       379.6448
## ******************************************************************
```

What About A Difference Between Caffeine and Combo in Budding?

Step 1: Building the Reordered Budding Model

```
mod_WITHtotal_budding_reordered <- gamlss(num_asexual_buds ~ treatment *
    collection_day + re(random = ~1 | jar_code) + re(random = ~1 | total_num),
    family = NBII(), data = comparisons_data, control = gamlss.control(n.cyc = 250))
```

Step 3: Summarizing the Model

```
summary(mod_WITHtotal_budding_reordered)
```

```
## ******************************************************************
## Family:  c("NBII", "Negative Binomial type II")
##
## Call:  gamlss(formula = num_asexual_buds ~ treatment * collection_day +
##     re(random = ~1 | jar_code) + re(random = ~1 | total_num),
##     family = NBII(), data = comparisons_data, control = gamlss.control(n.cyc = 250))
##
##
## Fitting method: RS()
##
## -----------------------------------------------------------------
## Mu link function:  log
## Mu Coefficients:
##                                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)                      0.25060    0.32458   0.772   0.4438
## treatmentCombination             0.86654    0.41179   2.104   0.0405 *
## treatmentControl                -0.79621    0.54887  -1.451   0.1533
## treatmentEstradiol              -0.61863    0.49533  -1.249   0.2176
## collection_day                   0.52771    0.07878   6.698 1.95e-08 ***
## treatmentCombination:collection_day -0.18018  0.10232  -1.761   0.0845 .
## treatmentControl:collection_day  0.08091    0.13120   0.617   0.5403
## treatmentEstradiol:collection_day 0.14298   0.11771   1.215   0.2303
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## -----------------------------------------------------------------
## Sigma link function:  log
```

```
## Sigma Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -2.526      2.645  -0.955    0.344
##
## ---------------------------------------------------------------
## NOTE: Additive smoothing terms exist in the formulas:
##  i) Std. Error for smoothers are for the linear effect only.
## ii) Std. Error for the linear terms maybe are not accurate.
## ---------------------------------------------------------------
## No. of observations in the fit:  60
## Degrees of Freedom for the fit:  11.04057
##       Residual Deg. of Freedom:  48.95943
##                      at cycle:  8
##
## Global Deviance:     271.739
##             AIC:     293.8201
##             SBC:     316.9429
## ****************************************************************
```