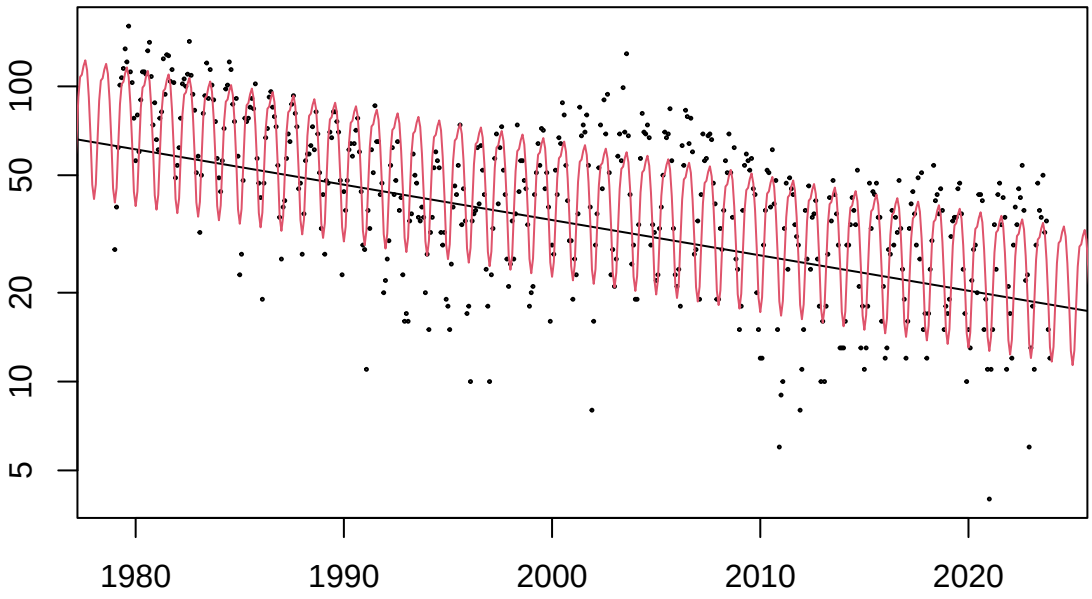
STA303 Assignment 2

Talia Fabregas Kasandra Tworzynaski

March 20, 2025

Motorcycle deaths




```
##         date killed serious slight dateInt logMonthDays month
## 1 1979-01-01     28     737   1677    3287     3.433987   Jan
## 2 1979-02-01     39     854   1914    3318     3.332205   Feb
## 3 1979-03-01     62    1302   2895    3346     3.433987   Mar
## 4 1979-04-01    101    1540   3406    3377     3.401197   Apr
## 5 1979-05-01    107    1774   4093    3407     3.433987   May
## 6 1979-06-01    115    1912   4441    3438     3.401197   Jun
```

1. Write down, in equations not R code, a generalized additive model suitable for this problem. Explain each of the parts of the model and give a rationale for them (i.e. "The response variable is Gamma distributed because the number of deaths must be positive"). (4 points)

   Do we have to check for overdispersion or is this already telling us it is poission?

   Number of motorcylcle deaths every month from 1980 to 2025

   An appropriate generalized additive model

   We chose a poisson because the response variable, the number of motorcycle deaths in month i, is a count variable.

   Covariate Xi is month

Beta is the fixed effect of month Motorcycle deaths

$$Y_i|U \sim Pois(D_i\lambda_i)G(\lambda_i) = X_i\beta + U(t_i)U(.) \sim IWP_2(\sigma)$$

Motorcycle deaths (analogous to w8 video slide)

$$Y_i|U \sim \text{NegBinom}(D_i, \lambda_i; \tau)g(\lambda_i) = X_i\beta + U(t_i)U(.) \sim \text{IWP}_2(\sigma)$$

- $E(Y_i|U) = \lambda_i$
- $sd(Y_i|U)/E(Y_i|U) = \tau$
- D_i = number of days in month i
- λ_i = number of motorcycle deaths per day in month i
- X_i = seasonality, where month is the factor

2. Show R code to fit the model using the `mcgv` package

```
motorcycleGAM <- gam(
  killed ~ offset(logMonthDays) + s(dateInt) + month,
  data=x,
  family=nb(link="log"),
  method="ML"
  )

# log offset on D_i
# smooth term on the data to capture long-term trend in num motorcycle deaths
# month, categorical variable, seasonal effects
```

3.

Heat

```
##           date killed serious slight dateInt logMonthDays month
## 100 1987-04-01     57    1090   2537    6299     3.401197   Apr
```

```
{r. echo=FALSE, message=FALSE, warning=FALSE} x$dateInt = as.integer(x$Date) x$yearFac
= factor(format(x$Date, "%Y")) # xSub = x[x$summer & !is.na(x$Max.Temp), ] # res1 =
gam(update.formula(Max.Temp ~ s(dateInt, pc = as.integer(as.Date("1990/7/1")), # +   k
= 100) + s(yearFac, bs = "re"), Pmisc::seasonalFormula(period = 365.25, # +   harmonics =
1:2, var = "dateInt")), data = xSub, method = "ML", # +   optimizer = "efs")
```

1.
2.
3.