In lecture 16, we looked at precipitation amounts in Madison County (at Morrisville station). We found that the Weibull distribution had a good fit to the monthly precipitation amounts.

We found that the MLEs for the Weibull distribution were

$$\hat{a} = 2.1871$$
$$\hat{\sigma} = 3.9683$$

and

$$-\mathcal{L}(\{\hat{a}, \hat{\sigma}\}|\mathbf{x}) = 2166.496$$

is the realized negative log-likelihood. Note this means that the log-likelihood is

$$\mathcal{L}(\{\hat{a}, \hat{\sigma}\}|\mathbf{x}) = -2166.496,$$

and the usual likelihood is

$$L(\{\hat{a}, \hat{\sigma}\}|\mathbf{x}) = e^{[\mathcal{L}(\{\hat{a},\hat{\sigma}\}|\mathbf{x})]} \approx= e^{-2166.496},$$

which `R` cannot differentiate from 0.

1. Someone asked "why Weibull?" in class. That is, why wouldn't we use another right-skewed distribution like the Gamma (see Lecture 15), or the Log-Normal (see Lecture 17).

   (a) Compute the MLEs for these data using a Gamma distribution.
   `Solution:` The MLEs for these data using a Gamma distribution are

   $$\hat{a} = 4.17$$
   $$\hat{\beta} = 1.19$$

   $$\mathcal{L}(\{\hat{a}, \hat{\beta}\}|\mathbf{x}) = -2151.149$$

   (b) Compute the MLEs for these data using the Log-Normal distribution.
   `Solution:` The MLEs for these data using a Log-Normal distribution are

   $$\hat{\mu} = 1.131$$
   $$\hat{\sigma} = 0.533$$

   $$\mathcal{L}(\{\hat{\mu}, \hat{\sigma}\}|\mathbf{x}) = -2204.201$$

   (c) Compute the likelihood ratio to compare the Weibull and the Gamma distribution. Which has a better fit according to the likelhiood ratio?

   $$Q = \frac{L(\{\hat{a}, \hat{\sigma}\}|\mathbf{x})}{L(\{\hat{\alpha}, \hat{\beta}\}|\mathbf{x})} = e^{\left[\mathcal{L}(\{\hat{a},\hat{\sigma}\}|\mathbf{x}) - \mathcal{L}(\{\hat{\alpha},\hat{\beta}\}|\mathbf{x})\right]}$$

   `Solution:` According to the likelihood ratio, the Weibull distribution has a better fit since the ratio has a value > 1, suggesting the top model (Weibull) as the better fit.

   (d) Compute the likelihood ratio to compare the Weibull and the Log-Normal distribution. Which has a better fit according to the likelihood ratio?

   $$Q = \frac{L(\{\hat{a}, \hat{\sigma}\}|\mathbf{x})}{L(\{\hat{\mu}, \hat{\sigma}\}|\mathbf{x})} = e^{[\mathcal{L}(\{\hat{a},\hat{\sigma}\}|\mathbf{x}) - \mathcal{L}(\{\hat{\mu},\hat{\sigma}\}|\mathbf{x})]}$$

   `Solution:` According to the likelihood ratio, the Log-Normal distribution has a better fit since the ratio has a value < 1, suggesting the bottom model (Log-Normal) as the better fit.

(e) Compute the likelihood ratio to compare the Gamma and the Log-Normal distribution. Which has a better fit according to the likelhiood ratio?

$$Q = \frac{L(\{\hat{\alpha}, \hat{\beta}\}|\mathbf{x})}{L(\{\hat{\mu}, \hat{\sigma}\}|\mathbf{x})} = e^{\left[\mathcal{L}(\{\hat{\alpha}, \hat{\beta}\}|\mathbf{x}) - \mathcal{L}(\{\hat{\mu}, \hat{\sigma}\}|\mathbf{x})\right]}$$

`Solution:` According to the likelihood ratio, the Log-Normal distribution has a better fit since the ratio has a value < 1, suggesting the bottom model (Log-Normal) as the better fit

`Code for Question 1:`

```r
rain.data = read_csv("agacis.csv") #Load the data set

## Rows:  115 Columns:  14
## -- Column specification --------------------------------------------------------
## Delimiter:  ","
## chr (13):  Jan, Feb, Mar, Apr, May, Jun, Jul, Aug, Sep, Oct, Nov, Dec, Annual
## dbl  (1):  Year
##
## 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.

rain.data <- rain.data |>
  select(-Annual) |>                           # Remove annual column
  pivot_longer(cols = c(Jan, Feb, Mar, Apr,    # pivot the column data into one col
                        May, Jun, Jul, Aug,
                        Sep, Oct, Nov, Dec),
               values_to = "Precipitation",    # store the values in Precipitation
               names_to = "Month") |>          # store the months in Month
  mutate(Precipitation = case_when(Precipitation == "M" ~ NA_character_,
                                   TRUE                 ~ Precipitation))|>
  mutate(Precipitation = as.numeric(Precipitation))

MLE.Weibull <- function(par,    #parameters arguement
                        data,   #dataframe arguement
                        neg=F){ #negative arguement
  #FUNCTION PURPOSE:
  #Use Maximum Log Likelihood to estimate parameters for a Weibull distribution for
  #a given data set
  a <- par[1]     #Alpha parameter
  sigma <- par[2] #Sigma parameter

  #Calculate the log likelihood for given parameters
  ll.like <- sum(log(dweibull(x = data, shape = a, scale = sigma)), na.rm=T)

  return(ifelse(neg, -ll.like, ll.like)) #If neg is True, return the result as a negative
}

MLE.Weibull.data <- optim(fn = MLE.Weibull, #Using optim, estimate the parameters
                          par = c(1,1),     #at which the log likelihood is at it's min
                          data = rain.data$Precipitation,
                          neg=T)            #(i.e. finds the parameters for the data's distribution)

MLE.Gamma <- function(data,          #dataframe argument
                      para,          #parameters argument
                      neg = FALSE) { #negative argument
  #FUNCTION PURPOSE:
  #Use Maximum Log Likelihood to estimate parameters for a Gamma distribution for
  #a given data set
  alpha <- para[1] #Alpha parameter
  beta <- para[2]  #Beta parameter

  #Calculate the log likelihood for given parameters
  ll.like <- sum(log(dgamma(data, shape = alpha, rate = beta)), na.rm = TRUE)

  return(ifelse(neg, -ll.like, ll.like)) #If neg is True, return the result as a negative
}

MLE.Gamma.data <- optim(fn = MLE.Gamma, #Using optim, estimate the parameters
                        par = c(1,1),   #at which the log likelihood is at it's min
                        data = rain.data$Precipitation,
                        neg = TRUE)     #(i.e. finds the parameters for the data's distribution)

MLE.LogNorm <- function(data,          #dataframe argument
                        para,          #parameters argument
                        neg = FALSE) { #negative argument
```

```r
  #FUNCTION PURPOSE:
  #Use Maximum Log Likelihood to estimate parameters for a Log Normal distribution for
  #a given data set
  mu <- para[1]     #Mu parameter
  sigma <- para[2]  #Sigma parameter

  #Calculate the log likelihood for given parameters
  ll.like <- sum(log(dlnorm(data, meanlog = mu, sdlog = sigma)), na.rm = TRUE)

  return(ifelse(neg, -ll.like, ll.like)) #If neg is True, return the result as a negative
}

MLE.LogNorm.data <- optim(fn = MLE.LogNorm,  #Using optim, estimate the parameters
                          par = c(1,1),      #at which the log likelihood is at it's min
                          data = rain.data$Precipitation,
                          neg = TRUE)        #(i.e. finds the parameters for the data's distribution)

LR1 <- MLE.Weibull.data$value/MLE.Gamma.data$value   #Calculates the Likelihood ratio for Weibull and Gamma distribution
LR2 <- MLE.Weibull.data$value/MLE.LogNorm.data$value #Calculates the Likelihood ratio for Weibull and Log Normal distribution
LR3 <- MLE.Gamma.data$value/MLE.LogNorm.data$value   #Calculates the Likelihood ratio for Gamma and Log Normal distribution
```

2. Optional Coding Challenge. Choose the "best" distribution and refit the model by season.

   (a) Fit the Distribution for Winter (December-February).
   `Solution:` The best fit distribution for Winter is a Weibull distribution.

   (b) Fit the Distribution for Spring (March-May).
   `Solution:` The best fit distribution for Spring is a Lognormal distribution.

   (c) Fit the Distribution for Summer (June-August).
   `Solution:` The best fit distribution for Summer is a Lognormal distribution.

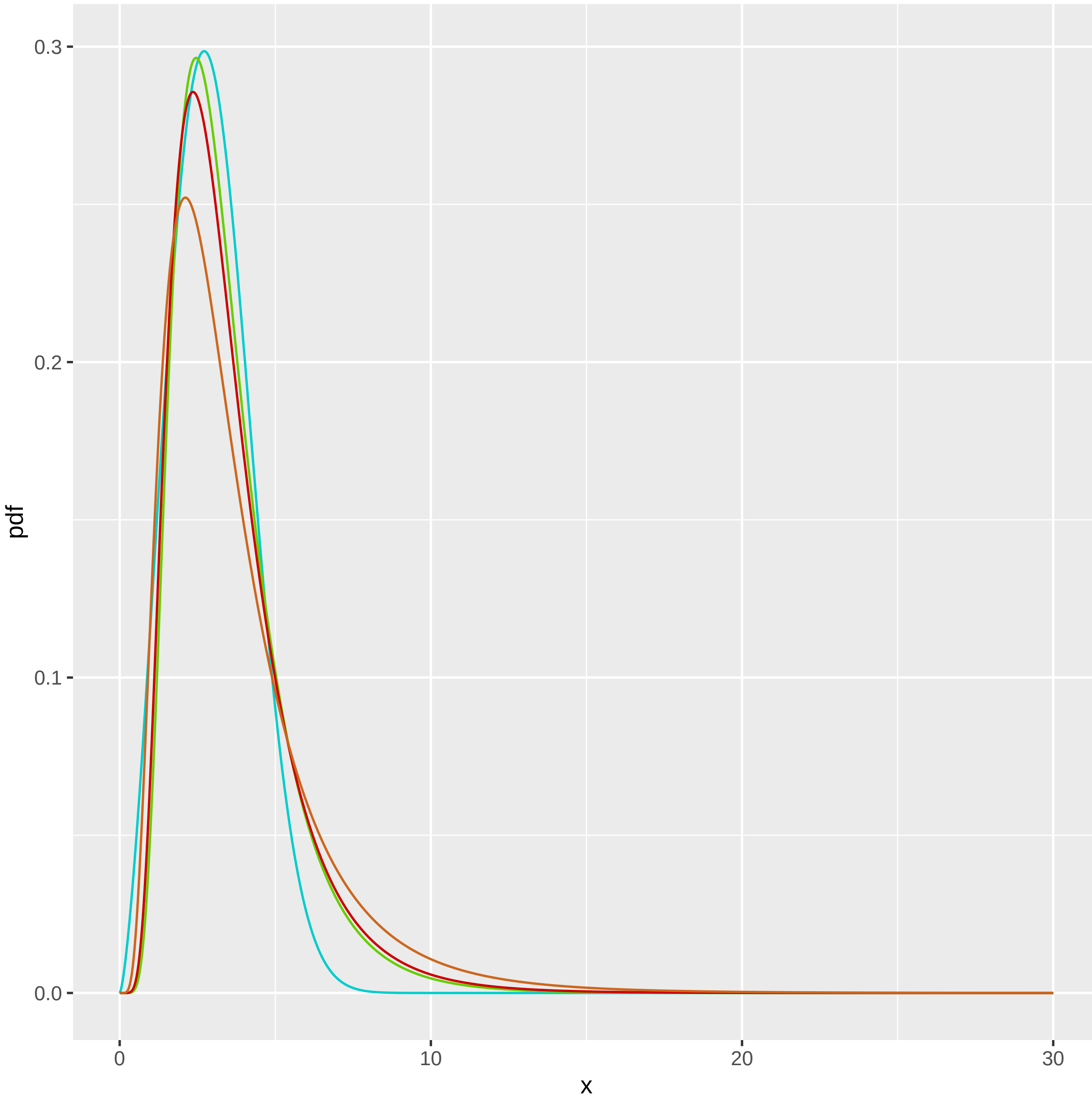
   (d) Fit the Distribution for Fall (September-November).
   `Solution:` The best fit distribution for Fall is a Lognormal distribution.

   (e) Plot the four distributions in one plot using `cyan3` for Winter, `chartreuse3` for Spring, `red3` for Summer, and `chocolate3` for Fall. Note any similarities/differences you observe across the seasons.
   `Solution:` For the most part all the distributions seem to be relatively similar to eachother, with some being a little more right skewed and platykurtic than the others. The fall distribution differs a little more than all the others (is more platykurtic).

`Code for Question 2:`

```r
winter.data <- rain.data |>
  filter(Month %in% c("Dec", "Jan", "Feb")) #Filter only the winter months

spring.data <- rain.data |>
  filter(Month %in% c("Mar", "Apr", "May")) #Filter only the spring moneths

summer.data <- rain.data |>
  filter(Month %in% c("Jun", "Jul", "Aug")) #Filter only the summer months

fall.data <- rain.data |>
  filter(Month %in% c("Sep", "Oct", "Nov")) #Filter only the fall months

data.list <- list(winter = winter.data, #Place all the filtered data sets into a list
                  spring = spring.data,
                  summer = summer.data,
                  fall = fall.data)
best.fit.results = c() #Will store the best distribution for each season

for(data in data.list) { #For each season
  MLE.Weibull.data <- optim(fn = MLE.Weibull, #Gather the Log Likelihood estimate
                            par = c(1,1),     #using optim
                            data = data$Precipitation,
                            neg=T)
  MLE.Gamma.data <- optim(fn = MLE.Gamma,     #Gather the Log Likelihood estimate
                          par = c(1,1),       #using optim
                          data = data$Precipitation,
                          neg = TRUE)
  MLE.LogNorm.data <- optim(fn = MLE.LogNorm, #Gather the Log Likelihood estimate
                            par = c(1,1),     #using optim
```

```r
                            data = data$Precipitation,
                            neg = TRUE)

  LR1 <- MLE.Weibull.data$value/MLE.Gamma.data$value   #Calculates the Likelihood ratio for Weibull and Gamma distribution
  LR2 <- MLE.Weibull.data$value/MLE.LogNorm.data$value #Calculates the Likelihood ratio for Weibull and Log Normal distribution
  LR3 <- MLE.Gamma.data$value/MLE.LogNorm.data$value   #Calculates the Likelihood ratio for Gamma and Log Normal distribution

  if(LR1 > 1 && LR2 > 1){ #If Weibull is favored between the Gamma and Log Normal
    best.fit <- "Weibull" #Assign best fit to Weibull
  } else if(LR1 < 1 && LR3 > 1){ #If Gamma is favored between the Weibull and Log Normal
    best.fit <- "Gamma"          #Assign best fit to Gamma
  } else if(LR2 < 1 && LR3 < 1) { #If Log Normal is favored between Weibull and Gamma
    best.fit <- "Log Normal"      #Assign best fit to Log Normal
  } else {
    best.fit <- "Tie" #If there are two distributions with equal fits, assign best fit to be a tie
  }

  best.fit.results <- c(best.fit.results, best.fit) #Store the best fit distributin for each season
}

winter.plot.data <- tibble(x = seq(0, 30, length.out = 1000)) %>% #Obtains data points to plot
  mutate(pdf = dweibull(x, shape = 2.47, scale = 3.35))           #Winter data distribution

spring.plot.data <- tibble(x = seq(0, 30, length.out = 1000)) %>% #Obtains data points to plot
  mutate(pdf = dlnorm(x = x, meanlog = 1.135, sdlog = 0.487))     #Spring data distribution

summer.plot.data <- tibble(x = seq(0, 30, length.out = 1000)) %>% #Obtains data points to plot
  mutate(pdf = dlnorm(x = x, meanlog = 1.126, sdlog = 0.518))     #Summer data distribution

fall.plot.data <- tibble(x = seq(0, 30, length.out = 1000)) %>%   #Obtains data points to plot
  mutate(pdf = dlnorm(x = x, meanlog = 1.130, sdlog = 0.619))     #Fall data distribution

ggplot() +
  geom_line(data = winter.plot.data, aes(x = x, y = pdf), color = "cyan3") + #draws winter distribution curve
  geom_line(data = spring.plot.data, aes(x = x, y = pdf), color = "chartreuse3") + #draws spring distribution curve
  geom_line(data = summer.plot.data, aes(x = x, y = pdf), color = "red3") + #draws summer distribution curve
  geom_line(data = fall.plot.data, aes(x = x, y = pdf), color = "chocolate3") #draws fall distribution curve
```



0.3
0.2
0.1
0.0
pdf
0
10
20
30
x