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.

```
library(tidyverse)

dat.precip <- read_csv(file = "agacis.csv")

dat.precip.long <- dat.precip |>
  dplyr::select(-Annual) |>
  pivot_longer(cols = c(Jan, Feb, Mar, Apr,
                        May, Jun, Jul, Aug,
                        Sep, Oct, Nov, Dec),
               values_to = "Precipitation",
               names_to = "Month") |>
  mutate(Precipitation = case_when(Precipitation == "M" ~ NA_character_,
                                   TRUE                 ~ Precipitation))|>
  mutate(Precipitation = as.numeric(Precipitation))

#####################################
# a.) MLEs (Gamma)
#####################################
llgamma <- function(par, data, neg=F){
  alpha <- exp(par[1])
  beta <- exp(par[2])

  ll <- sum(log(dgamma(x=data, shape=alpha, scale=beta)), na.rm=T)

  return(ifelse(neg, -ll, ll))
}

MLE.gamma <- optim(fn = llgamma,
                   par = c(1,1),
                   data = dat.precip.long$Precipitation,
                   neg=T)

ll.gamma <- -MLE.gamma$value
(MLE.gamma$par <- exp(MLE.gamma$par))

## [1] 4.1761219 0.8405941
```

   (b) Compute the MLEs for these data using the Log-Normal distribution.

```
#####################################
# b.) MLEs (Lognormal)
#####################################
lllognormal <- function(par, data, neg=F){
  mu <- par[1]
  sigma <- exp(par[2])

  ll <- sum(log(dlnorm(x=data, meanlog=mu, sdlog=sigma)), na.rm=T)
```

```
  return(ifelse(neg, -ll, ll))
}

MLE.lognormal <- optim(fn = lllognormal,
                       par = c(0,1),
                       data = dat.precip.long$Precipitation,
                       neg=T)

ll.lognormal <- -MLE.lognormal$value
MLE.lognormal$par[2] <- exp(MLE.lognormal$par[2])
(MLE.lognormal$par)

## [1] 1.1313347 0.5333965
```

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

```
####################################
# c.) Weibull / Gamma
####################################

ll.weibull <- -2166.496
(Q.wg <- exp(ll.weibull - ll.gamma))

## [1] 2.161379e-07

# Q.wg < 1, therefore the Gamma distribution has a better fit than the Weibull distribution.
```

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

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

```
####################################
# d.) Weibull / Log-Normal
####################################

(Q.wl <- exp(ll.weibull - ll.lognormal))

## [1] 2.37065e+16

# Q.wl > 1, therefore the Weibull distribution has a better fit than the Log-Normal distribution.
```

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

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

```
####################################
# e.) Gamma / Log-Normal
####################################

(Q.gl <- exp(ll.gamma - ll.lognormal))

## [1] 1.096823e+23

# Q.gl > 1, therefore the Gamma distribution has a better fit than the Log-Normal distribution.
```

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

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

   (a) Fit the Distribution for Winter (December-February).

   (b) Fit the Distribution for Spring (March-May).

   (c) Fit the Distribution for Summer (June-August).

(d) Fit the Distribution for Fall (September-November).

(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.