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")

#data cleaning from lecture
dat.precip.long <- dat.precip |>
  dplyr::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))
#log likelihood function for Gamma
llgamma <- function(par, data, neg=F){
  alpha <- par[1]
  beta <- par[2]

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

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

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

## [1] 4.174581 1.189099

(gamma.MLEs$par <- exp(gamma.MLEs$par)) # transform parameters to usual

## [1] 65.012618  3.284122
```

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

```
#log likelihood function for Log normal
lllognorm <- function(par, data, neg=F){
mu <- par[1]
sigma <- par[2]

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

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

```r
lognorm.MLEs <- optim(fn = lllognorm,
           par = c(1,1),
           data = dat.precip.long$Precipitation,
           neg=T)
lognorm.MLEs$par

## [1] 1.1312609 0.5333417

(lognorm.MLEs$par <- exp(lognorm.MLEs$par)) # transform parameters to usual

## [1] 3.099562 1.704619
```

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

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

```r
ll.weibull = -2166.496 #log-likelihood for Weibull
ll.gamma = -gamma.MLEs$value #log-likelihood for Gamma
(weibull.gamma = exp(ll.weibull - ll.gamma)) #likelihood ratio for Weibull and Gamma

## [1] 2.161318e-07
```

Because this ratio is very small, then the distribution in the numerator (Weibull) has a greater (magnitude) maximum likelihood value. So the Weibull distribution is a better fit to the data.

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

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

```r
ll.lognorm = -lognorm.MLEs$value #log-likelihood for Lognorm
(weibull.lognorm = exp(ll.weibull - ll.lognorm)) #likelihood ratio for Weibull and Lognorm

## [1] 2.370668e+16
```

Because this ratio is very large, then the distribution in the denominator (Log-Normal) has a greater (magnitude) maximum likelihood value. So the Log-Normal distribution is a better fit to the data.

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

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

```r
(gamma.lognorm = exp(ll.gamma - ll.lognorm)) #likelihood ratio for Gamma and Lognorm

## [1] 1.096862e+23
```

Because this ratio is very large, then the distribution in the denominator (Log-Normal) has a greater (magnitude) maximum likelihood value. So the Log-Normal distribution is a better fit to the data.

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.