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.

   The MLEs for the Gamma distribution are $\hat{\alpha} = 4.1742$, $\hat{\beta} = 0.8411$, and a negative log-likelihood $(-\mathcal{L}(\hat{\alpha}, \hat{\beta} \mid \mathbf{x})$ of 2151.149.

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

   The MLEs for the Log-Normal distribution are $\hat{\mu} = 1.1314$, $\hat{\sigma} = 0.5333$, and a negative log-likelihood $(-\mathcal{L}(\hat{\mu}, \hat{\sigma} \mid \mathbf{x})$ of 2204.201.

   (c) Compute the likelihood ratio to compare the Weibull and the Gamma distribution. Which has a better fit according to the likelihood 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
log.lik.weib <- -2166.496
log.lik.gamma <- -2151.149
(q <- exp(log.lik.weib - log.lik.gamma))

## [1] 2.162134e-07
```

   Since likelihood ratio, $Q < 1$, the Gamma distribution has a better fit for our 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?

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

```r
log.lik.weib <- -2166.496
log.lik.lognorm <- -2204.201
(q <- exp(log.lik.weib - log.lik.lognorm))

## [1] 2.371775e+16
```

   Since likelihood ratio $Q > 1$, the Weibull distribution has a better fit for our 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?

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

```
(q <- exp(log.lik.gamma - log.lik.lognorm))

## [1] 1.09696e+23
```

Since likelihood ratio $Q > 1$, the Gamma distribution has a better fit for our 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.

   Code for Q1:

```
library(tidyverse)
library(nleqslv)
data <- read_csv("agacis.csv")
dat.precip.long <- data |>                    # clean the data
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))
#####################################
# Maximum likelihood (Gamma)
#####################################
llgamma <- function(par, data, neg=F){
alpha <- par[1]
beta <- par[2]
loglik <- sum(log(dgamma(x=data, shape=alpha, scale = beta)), na.rm = T)
return(ifelse(neg, -loglik, loglik))
}
gamma.MLEs <- optim(fn = llgamma,
            par = c(1,1),
            data = dat.precip.long$Precipitation,
            neg=T)
#####################################
# Maximum likelihood (Log Normal)
#####################################
ll.lognorm <- function(par, data, neg=F){
mu <- par[1]
sigma <- par[2]
loglik <- sum(log(dlnorm(x=data, meanlog = mu, sdlog = sigma)), na.rm = T)
return(ifelse(neg, -loglik, loglik))
}
lognorm.MLEs <- optim(fn = ll.lognorm,
            par = c(0,1),
            data = dat.precip.long$Precipitation,
            neg=T)
```