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.

```
#####################################################################################################
"(1.a)  Compute the MLEs for these data using a Gamma distribution"

## [1] "(1.a)  Compute the MLEs for these data using a Gamma distribution"

llgamma <- function(data, par, neg=F){
  alpha <- par[1]
  beta <- par[2]

  loglik <- sum(log(dgamma(x=data, shape=alpha, rate=beta)))

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

mles <- optim(par = c(1,1),
               fn = llgamma,
               data=dat.precip.long$Precipitation,
               neg=T)
(alpha.hat.mle <- mles$par[1])

## [1] 4.174581

(beta.hat.mle <- mles$par[2])

## [1] 1.189099

(loglik_gamma <- -mles$value) # Calculate loglik_gamma

## [1] -2151.149

#####################################################################################################
```

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

```
#####################################################################################################
"(1.b)  Compute the MLEs for these data using the Log-Normal distribution"

## [1] "(1.b)  Compute the MLEs for these data using the Log-Normal distribution"

lllognormal <- function(data, par, neg=F) {
  meanlog <- par[1]
  sdlog <- par[2]

  loglik <- sum(log(dlnorm(x=data, meanlog=meanlog, sdlog=sdlog)))
```

```r
  return(ifelse(neg, -loglik, loglik))
}

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

(meanlog.hat.mle <- mles$par[1])

## [1] 1.131462

(sdlog.hat.mle <- mles$par[2])

## [1] 0.5333435

(loglik_lognormal <- -mles$value) # Calculate loglik_lognormal

## [1] -2204.201

#########################################################################################################
```

(c) Compute the likelihood ratio to compare the Weibull and the Gamma distribution. Which has a better fit according to the likeihood ratio? **Since the likelihood ratio comparing the Weibull and the Gamma distribution is 2.162312e-07 which is less than one, it indicates that the Gamma distribution is a better fit than the Weibull distribution.**

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

```r
##########################################
#Maximum Likelihood of Weibull Distibution (Taken from rcode-lecture16.R)

llweibull <- function(par, data, neg=F){
  # a <- par[1]
  # sigma <- par[2]
  a <- exp(par[1]) # go from (-inf,inf) to (0,inf)
  sigma <- exp(par[2]) # go from (-inf,inf) to (0,inf)

  ll <- sum(log(dweibull(x=data, shape=a, scale=sigma)), na.rm=T)

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

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

(MLEs$par <- exp(MLEs$par)) # transform

## [1] 2.187091 3.968269

(loglik_weibull <- -MLEs$value) # Calculate loglik_weibull

## [1] -2166.496

#########################################################################################################
# Compute the likelihood ratio to compare the Weibull and the Gamma distribution

(Q.WG = exp(loglik_weibull - loglik_gamma))

## [1] 2.162312e-07

#########################################################################################################
```

(d) Compute the likelihood ratio to compare the Weibull and the Log-Normal distribution. Which has a better fit according to the likelihood ratio? **Since the likelihood ratio comparing the Weibull and the Log-Normal distribution is 2.371759e+16 which is greater than one, it indicates that the Weibull distribution is a better fit than the Log-Normal distribution.**

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

```r
##########################################################################################################
# "(1.d)  Compute the likelihood ratio to compare the Weibull and the Log-Normal distribution.
(Q.WLG = exp(loglik_weibull - loglik_lognormal))
## [1] 2.371759e+16
##########################################################################################################
```

(e) Compute the likelihood ratio to compare the Gamma and the Log-Normal distribution. Which has a better fit according to the likelihood ratio? **Since the likelihood ratio comparing the Gamma and the Log-Normal distribution is 1.096862e+23 which is greater than one, it indicates that the Gamma distribution is a better fit than the Log-Normal distribution.**

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

```r
##########################################################################################################
# "(1.e)  Compute the likelihood ratio to compare the Gamma and the Log-Normal distribution.
(Q.GLG = exp(loglik_gamma - loglik_lognormal))
## [1] 1.096862e+23
##########################################################################################################
```

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.