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.

```
## # A tibble: 1 x 6
##    mean    sd   min   max  skew  kurt
##   <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1  3.51  1.69   0.1  12.3  1.02  1.92
```

```
###########################################################################
# a) Compute the MLEs for these data using a Gamma distribution
###########################################################################
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))
}

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

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

## [1] 4.1761219 0.8405941
```

   To estimate the parameters of the Gamma distribution, we used maximum likelihood estimation (MLE). The estimated parameters are:

$$\hat{\alpha} = 4.176, \quad \hat{\beta} = 0.841$$

   We followed the same MLE process as in class, using `optim()` to find the values of $\hat{\alpha}$ and $\hat{\beta}$ that maximize the likelihood function. Of particular note is our transformation of α and β. Since the shape (α) and scale(β) parameters of the Gamma distribution must be positive, we transformed them as $\alpha = \exp(\alpha), \quad \beta = \exp(\beta)$. This ensures the `optim()` algorithm operates over the entire real number space while keeping the parameters strictly positive.

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

```r
############################################################################
# b) Compute the MLEs for these data using the Log-Normal distribution
############################################################################

lllognorm <- 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))
}

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

(MLEs_lognorm$par[1])

## [1] 1.131308

(exp(MLEs_lognorm$par[2]))

## [1] 0.5333176
```

To estimate the parameters of the Log-Normal distribution, we used maximum likelihood estimation (MLE). The estimated parameters are:

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

Since the standard deviation parameter (σ) must be positive we applied the transformation as above. However, the mean parameter μ does not have the same restrictions, so there was no need to transform that parameter value. Once again, we followed the same MLE procedures.

(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
############################################################################
# c)  Compute the likelihood ratio to compare the Weibull and the Gamma dist
############################################################################

# compute the log-likelihood values at MLEs
ll_weibull <- -2166.496
ll_gamma <- -MLEs_gamma$value

#compute the likelihood ratio
(ratio_w_g <- exp(ll_weibull-ll_gamma))

## [1] 2.161379e-07
```

To compare the Weibull and Gamma distributions, we computed the likelihood ratio, which gave us the value of $Q = 2.16 \times 10^{-7}$. From this, we determine that the Gamma distribution provides a better fit than the Weibull distribution, as the likelihood ratio is much smaller than 1. This suggests that the Gamma model captures the precipitation data more effectively.

(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
############################################################################
# d)  Compute the likelihood ratio to compare the Weibull and the Lognormal dist
############################################################################
```

```r
# compute the log-likelihood values at MLEs
ll_lognorm <- -MLEs_lognorm$value

#compute the likelihood ratio
(ratio_w_ln <- exp(ll_weibull-ll_lognorm))

## [1] 2.370639e+16
```

Similarly, we compared the Weibull and Log-Normal distributions using the likelihood ratio, which gave us the value of $Q = 2.37 \times 10^{16}$. This result is extremely large, indicating that the Weibull distribution fits the data far better than the Log-Normal distribution. This suggests that the Log-Normal model is a poor choice compared to the Weibull model.

(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
###########################################################################
# e)  Compute the likelihood ratio to compare the Gamma and the Lognormal dist
###########################################################################
(ratio_g_ln <- exp(ll_gamma-ll_lognorm))

## [1] 1.096818e+23
```

Finally, we computed the likelihood ratio comparing the Gamma and Log-Normal distributions, which gave us the value of $Q = 1.10 \times 10^{23}$, which is overwhelmingly large. This suggests the Gamma distribution provides a much better fit than the Log-Normal distribtion, reinforcing the conclusion that the Log-Normal model is the weakest among the three considered distributions.

Overall, among the three distributions, the Gamma distribution is best fit for the data, as it outperforms both the Weibull and Log-Normal distributions. The second best choice is the Weibull distribution, and the Log-Normal distribution is the worst choice as its likelihood is significantly lower than both the Gamma and Weibull distributions.

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.