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.

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

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

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

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

## [1] 4.174581

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

## [1] 1.189099
```

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

```
lognorm <- function(data, par, neg=F){
 mean <- par[1]
 sd <- par[2]


 loglik <- sum(log(dlnorm(x=data, meanlog = mean, sdlog = sd)), na.rm = T)

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

```r
mles.ln <- optim(par = c(1,0.5),
            fn = lognorm,
            data= dat.precip.long$Precipitation,
            neg=T)
(mu <- mles.ln$par[1])
```

```
## [1] 1.131456
```

```r
(sigma <- mles.ln$par[2])
```

```
## [1] 0.5333391
```

(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
neg.llweibull <- -2166.496
neg.llgamma <- -(mles.gamma$value)
(weibull.gamma.ratio <- exp(neg.llweibull-neg.llgamma))
```

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

Since the ratio is extremely small, the likelihood for the weibull distribution is smaller than the gamma and suggests that the gamma distribution is a better fit to this data set.

(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
neg.llweibull <- -2166.496
neg.lognorm <- -(mles.ln$value)
(weibull.lognorm.ratio <- exp(neg.llweibull-neg.lognorm))
```

```
## [1] 2.370663e+16
```

Since the ratio is extremely large, the likelihood for the weibull distribution is greater than the log-normal and suggests that the weibull distribution is a better fit to this data set.

(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.ratio <- exp(neg.llgamma-neg.lognorm))
```

```
## [1] 1.09686e+23
```

Since the ratio is extremely large, the likelihood for the gamma distribution is greater than the log-normal and suggests that the gamma distribution is a better fit to this data set.

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.