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(par, data, neg=F){
  alpha <- exp(par[1]) # need alpha greater than 0 bc alpha E R+
  beta <- exp(par[2]) # need beta greater than 0 bc beta E R+

  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)

(hat.alpha <- exp(MLEs.gamma$par[1]))

## [1] 4.176122

(hat.beta <- exp(MLEs.gamma$par[2]))

## [1] 0.8405941
```

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

```
lllognormal <- function(par, data, neg=F){
  mu <- par[1]
  sigma <- exp(par[2]) # need sigma greater than 0 bc sigma E R+

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

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

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

(hat.mu <- MLEs.lnorm$par[1])

## [1] 1.131308

(hat.sigma <- exp(MLEs.lnorm$par[2]))

## [1] 0.5333176
```

(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
# Weibull vs Gamma
weibull.loglikelihood <- -2166.496
gamma.loglikelihood <- -MLEs.gamma$value

(Q <- exp(weibull.loglikelihood-gamma.loglikelihood))

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

The likelihood ratio of the Weibull and the Gamma distribution is `Q = 2.161379e-07` which is less than 1. This indicates that the Gamma distribution is a better fit according to the likelihood ratio.

(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
# Weibull vs Log-Normal
weibull.loglikelihood <- -2166.496
lnorm.loglikelihood <- -MLEs.lnorm$value # needs to be negative to
                                         # get the log-likelihood

(Q2 <- exp(weibull.loglikelihood-lnorm.loglikelihood))

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

The likelihood ratio of the Weibull and the Log-Normal distribution is `Q = 2.370639e+16` which is greater than 1. This indicates that the Weibull distribution is a better fit according to the likelihood ratio.

(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 vs Log-Normal
gamma.loglikelihood <- -MLEs.gamma$value
lnorm.loglikelihood <- -MLEs.lnorm$value

(Q3 <- exp(gamma.loglikelihood - lnorm.loglikelihood))

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

The likelihood ratio of the Gamma and the Log-Normal distribution is `Q = 1.096818e+23` which is greater than 1. This indicates that the Gamma distribution is a better fit according to the likelihood ratio.

From our analysis, we can see that the overall takeaway of what right-skewed distribution we could use to fit the monthly precipitation amounts in Madison County is that Gamma is the best fit, Weibull is the second best fit, and Log-Normal is the worst fit according to the likelihood ratios.

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.