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.

```
data.precip = read_csv("data.csv")

mlegamma <- 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 = mlegamma,
           data=data.precip$Precipitation,
           neg=T)
(alpha.hat.mle <- mles$par[1])

## [1] 4.174581

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

## [1] 1.189099

#LOG LIKELIHOOD
(sum(log(dgamma(x=data.precip$Precipitation, shape = alpha.hat.mle, rate = beta.hat.mle))))

## [1] -2151.149
```

$$\hat{\alpha} = 4.1746$$
$$\hat{\beta} = 1.1891$$

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

```
data.precip = read_csv("data.csv")

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

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

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

```
}

mles <- optim(par = c(1,1),
          fn = mlelognormal,
          data=data.precip$Precipitation,
          neg=T)
(meanlog <- mles$par[1])

## [1] 1.131261

(sdlog <- mles$par[2])

## [1] 0.5333417

#LOG LIKELIHOOD
(sum(log(dlnorm(x=data.precip$Precipitation, meanlog=meanlog, sdlog=sdlog))))

## [1] -2204.201
```

$$\hat{\mu} = 1.1312$$
$$\hat{\sigma} = 0.5333$$

(c) Compute the likelihood ratio to compare the Weibull and the Gamma distribution. Which has a better fit according to the likelhiood ratio?

$$Q = \frac{L(\{\hat{a}, \hat{\sigma}\}|\mathbf{x})}{L(\{\hat{\alpha}, \hat{\beta}\}|\mathbf{x})} = e^{\left[\mathcal{L}(\{\hat{a},\hat{\sigma}\}|\mathbf{x})-\mathcal{L}(\{\hat{\alpha},\hat{\beta}\}|\mathbf{x})\right]} \approx e^{-15.34738}$$

Explanation: Since the Q-value is less than one then we know that the Gamma distribution has a better fit according to the 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?

$$Q = \frac{L(\{\hat{a}, \hat{\sigma}\}|\mathbf{x})}{L(\{\hat{\mu}, \hat{\sigma}\}|\mathbf{x})} = e^{\left[\mathcal{L}(\{\hat{a},\hat{\sigma}\}|\mathbf{x})-\mathcal{L}(\{\hat{\mu},\hat{\sigma}\}|\mathbf{x})\right]} \approx e^{37.70453}$$

Explanation: Since the Q-value is greater than one then we know that the Weibull distribution has a better fit according to the 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?

$$Q = \frac{L(\{\hat{\alpha}, \hat{\beta}\}|\mathbf{x})}{L(\{\hat{\mu}, \hat{\sigma}\}|\mathbf{x})} = e^{\left[\mathcal{L}(\{\hat{\alpha},\hat{\beta}\}|\mathbf{x})-\mathcal{L}(\{\hat{\mu},\hat{\sigma}\}|\mathbf{x})\right]} \approx e^{53.05191}$$

Explanation: Since the Q-value is greater than one then we know that the Gamma distribution has a better fit according to the ratio.

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.

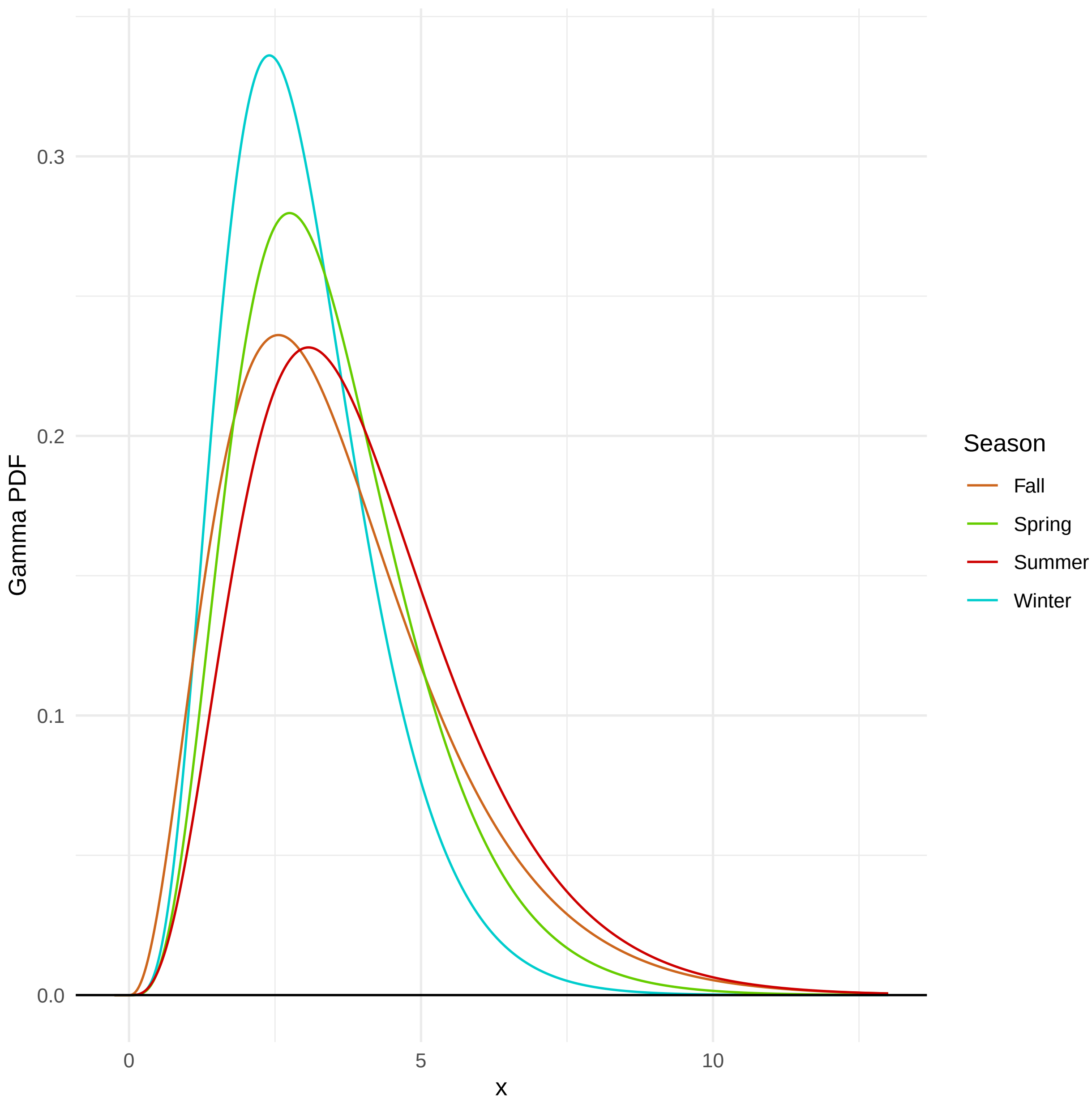

Figure 1: Gamma Dist. for Each Season