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 this data using a Gamma distribution.
   Solution: The MLEs for the data using a Gamma distribution are:

   $$\hat{\alpha} = 4.1761219$$

   $$\hat{\beta} = 0.8405941.$$

   (b) Compute the MLEs for these data using the Log-Normal distribution.
   Solution: The MLEs for this data using a Log-Normal distribution are:

   $$\hat{\mu} = 1.1313091$$

   $$\hat{\sigma} = 0.5333472.$$

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

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

   Solution: The likelihood ratio of the Weibull and Gamma distribution is $Q = 2.1623728 \times 10^{-7}$. Because $Q < 1$, the denominator is larger than the numerator, meaning that the likelihood of the denominator is larger, meaning it is a better fit. This means that the Gamma distribution is a better fit.

   (d) Compute the likelihood ratio to compare the Weibull and the Log-Normal distribution. Which has a better fit according to the likelihood ratio?

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

   Solution: The likelihood ratio of the Weibull and Log-Normal distribution is $Q = 2.371723 \times 10^{16}$. Because $Q > 1$, the numerator is larger than the denominator, meaning that the likelihood of the numerator is larger, meaning it is a better fit. This means that the Weibull distribution is a better fit.

(e) Compute the likelihood ratio to compare the Gamma and the Log-Normal distribution. Which has a better fit according to the likelhiood ratio?

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

Solution: The likelihood ratio of the Gamma and Log-Normal distribution is $Q = 1.0968151 \times 10^{23}$. Because $Q > 1$, the numerator is larger than the denominator, meaning that the likelihood of the numerator is larger, meaning it is a better fit. This means that the Gamma distribution is a better fit.

Overall, the Gamma distribution is the best fit, followed by the Weibull distribution and the Log-Normal distribution. Below is a plot (Figure 1) showing the estimated PDFs from the MLEs for each distribution. Notice how the Weibull PDF seems to underestimate the data a little bit, and the Log-Normal PDF overestimates the data a bit, with the Gamma PDF showing the best fit.

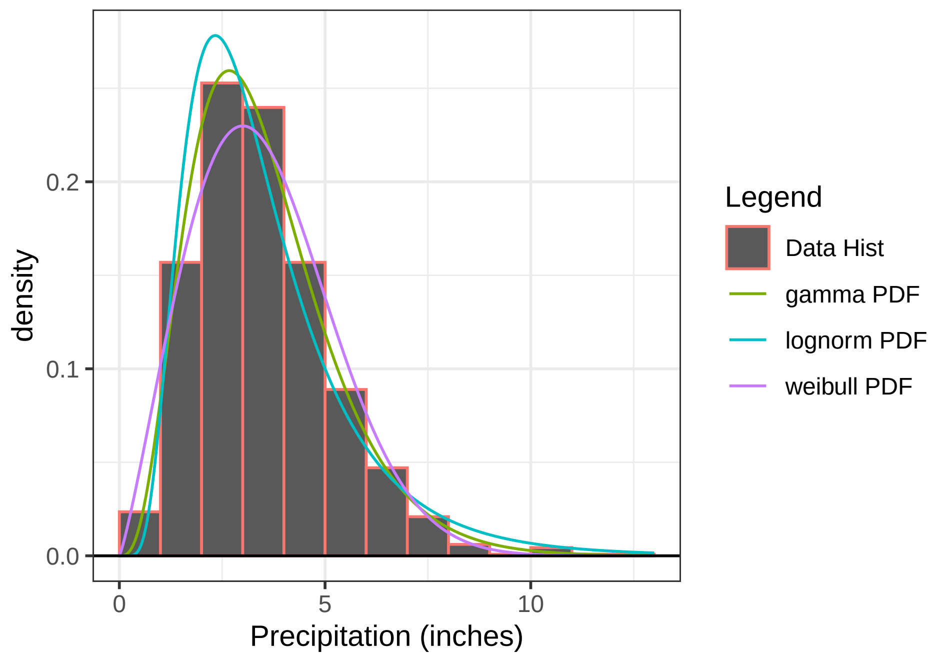

Figure 1: A plot with the histogram of the data, and the MLE estimates for the PDF for the Gamma, Log-Normal, and Weibull distributions.

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.