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.

   Solution: The MLEs using a Gamma distribution are $\alpha = 4.174581, \beta = 1.189099$.

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

   Solution: The MLEs for the data using the Log-Normal distribution are $\mu = 1.189099, \sigma = 0.5333435$.

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

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

   Solution: After using `optim()` to calculate the MLEs for the Weibull, Gamma, and Log-Normal distributions, we can access the likelihood statistic via:

```
weibull.loglike <- -MLEs$value
gamma.loglike <- -gamma.mles$value
lognorm.loglike <-  -lognorm.mles$value

Q <- exp(weibull.loglike - gamma.loglike)
```

   We then use the given formula for Q to calculate the likelihood ratio of $2.162312e-07$. This is approximately zero, so the gamma and Weibull fits are essentially the same here. Note that we have to add the negative in, as `optim()` outputs the positive value, aka. the realized negative log-likelihood. So, we convert to log-likelihood via the negative.

   (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: We repeat the same process to compare between the Weibull and Log-Normal distribution to get $Q = 2.371759e+16$. This is greater than zero and means that the Weibull distribution is a much better fit than the Log-Normal distribution.

(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: Again, we repeat the same process as for Weibull and Gamma, resulting in $Q = 1.096862e + 23$, which indicates that the Gamma distribution is a better fit than the Log-Normal distribution.

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.