Exercise 2

Data Analysis

March 20, 2024

Use again the `flights` data set as in the last exercise.

1. Show `air_time` in a histogram with relative frequencies, and plot on top of the histogram the density estimate. What do you see?

2. Do the same as in 1. for the variable `distance`.

3. Compare the empirical distribution function for `air_time` and `distance` with the function `ecdf()`. Try to plot the two functions on top of each other. Since the scale is very different, you should first try to normalize the variables appropriately. Which conclusions can you draw?

4. Use the function `qqPlot()` of the package `car`, and check if the number of flights is uniformly distributed over the months (1–12). Do the same for the days (1–31).

5. Similar as before: Check if the number of flights to the different destinations follows a normal, log-normal, or exponential distribution.

6. Compare the departure times of the carriers `AA` and `WN` in a QQ-plot by usimg the function `qqplot()`. What can you conclude?

7. Do the same as in 6. "by hand". I.e., plot the quantiles of the departure time of carrier `AA` versus those of carrier `WN`.
   Hint: Use the function `quantile()`

8. Show the density estimate of the arrival delay. Cut the x-range in the plot to just focus on the main part of the distribution. Show on top of that by differently colored lines the density estimates for the single main carriers (with at least 1000 flights). What can you conclude?

9. Do the same as in 8., but with QQ-plots (comparing with quantiles of the normal distribution).