# **Data Description and Scenario**

Rose-Hulman rarely has snow days since the Terre Haute, IN area does not experience extreme amounts of snow often. However, in the winter of 2021, the school had to call a snow day due to travel restrictions in Indiana that had resulted from a larger than usual snow event. There are several larger U.S. cities at similar latitudes as Terre Haute that might be hypothesized to experience similar climate conditions. In this data analysis, you’ll consider snowfall observed in Saint Louis, MO, Louisville, KY, Indianapolis, IN, and Cincinnati, OH (locations relative to Terre Haute shown in the map below).

Map

Description automatically generated

Yearly total snowfall amounts for each of the four locations (City) in the years of 2010 – 2020 (Year) were obtained. The Days variable contains the number of days with 1 or more inches that year and the snowfall totals (Snowfall.inches) are given in inches.

(Original Data Source: https://www.currentresults.com)

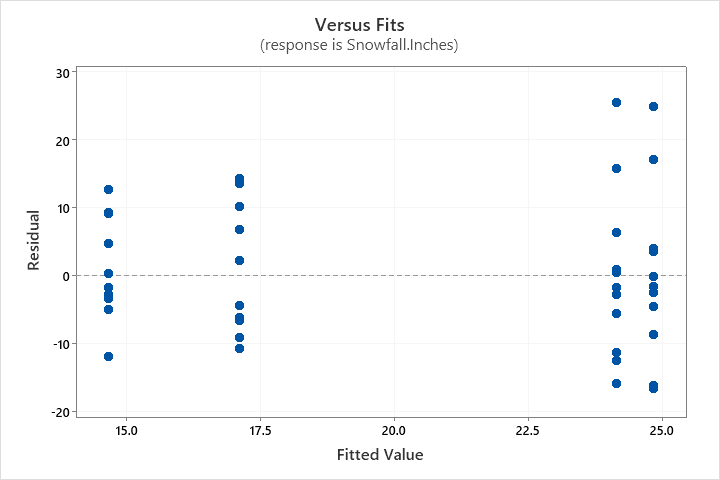
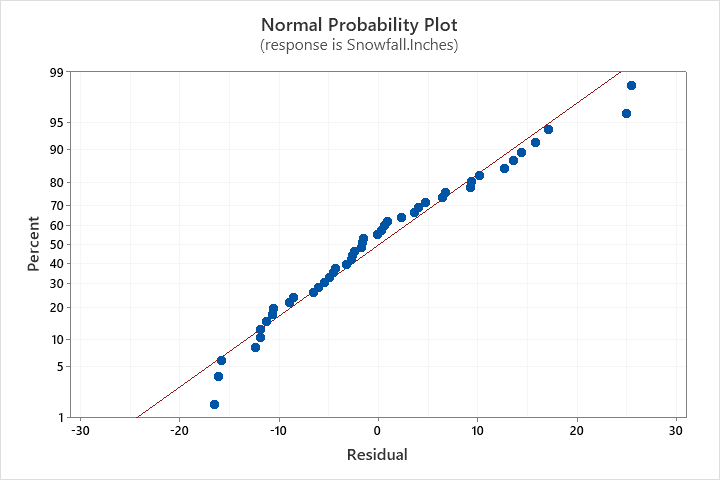
Let’s define parameters as follows.

µk: average yearly snowfall total at location k over all years, in inches

k=S (Saint Louis), I (Indianapolis), L (Louisville), C (Cincinnati)

H0: µS = µI = µL = µC H1: at least one µk differs

Residual plots from the one-way ANOVA model fit are provided on the next page. Notice that, even though we have yearly information, the individual observations/cases within the data aren’t truly ordered so a residuals vs. order plot cannot be constructed.

# **Questions**

1. Is it reasonable to assume that errors are independent? Briefly justify.
2. Is it reasonable to assume that errors have mean 0? Briefly justify.
3. Is it reasonable to assume that errors have constant variance? Briefly justify.
4. Is it reasonable to assume that errors come from a normal distribution? Briefly justify.

# **Solutions**

1. No. Even though we do not have a residuals vs. order plot, there is still yearly information. Within any location, there may be dependence among observations/cases year-to-year that is not explained by estimating separate means by location. Moreover, within any year, there may be dependence among locations since weather patterns are often similar at this latitude in the Midwest. (Students living in the Midwest U.S. would be familiar with the second statement.)
2. Yes. This is automatically true in one-way ANOVA.
3. Yes. There does not appear to be a drastic difference in spreads among the vertical lines of residuals in the vs. fits plot. The difference in spreads is less than 2 times.
4. Yes. There is not a pattern away from the diagonal in the normal Q-Q plot of residuals. Although some points are further from the diagonal in the tails of the plot, this is typical to observe, even in data from the normal distribution.